US009342146B2

(12) **United States Patent**
Bychkov et al.

(10) **Patent No.: US 9,342,146 B2**
(45) **Date of Patent: May 17, 2016**

(54) **POINTING-BASED DISPLAY INTERACTION**

(71) Applicant: **PRIMESENSE LTD**, Tel Aviv (IL)

(72) Inventors: **Eyal Bychkov**, Hod Hasharon (IL); **Oren Brezner**, Rishon Lezion (IL); **Micha Galor**, Tel Aviv (IL); **Ofir Or**, Ramat Gan (IL); **Jonathan Pokrass**, Bat Yam (IL); **Amir Hoffnung**, Tel Aviv (IL); **Tamir Berliner**, Beit Hashmonay (IL)

(73) Assignee: **APPLE INC.**, Cupertino, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 152 days.

(21) Appl. No.: **13/960,823**

(22) Filed: **Aug. 7, 2013**

(65) **Prior Publication Data**

US 2013/0321271 A1 Dec. 5, 2013

Related U.S. Application Data

(63) Continuation of application No. PCT/IB2012/050577, filed on Feb. 9, 2012.

(60) Provisional application No. 61/440,877, filed on Feb. 9, 2011, provisional application No. 61/526,692, filed on Aug. 24, 2011, provisional application No. 61/538,867, filed on Sep. 25, 2011.

(51) **Int. Cl.**
G06F 3/01 (2006.01)

(52) **U.S. Cl.**
CPC ***G06F 3/013*** (2013.01); ***G06F 3/017*** (2013.01)

(58) **Field of Classification Search**
CPC G06F 3/013; G06F 3/017; G06F 3/0481; G06F 2203/04806; G06F 3/048; G06K 9/00597; G02B 27/0093; A61B 3/113
USPC 345/156, 661; 351/209; 382/103; 396/51
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,550,250 A 10/1985 Mueller et al.
4,789,921 A 12/1988 Aho
4,836,670 A 6/1989 Hutchinson
(Continued)

FOREIGN PATENT DOCUMENTS

WO 9935633 A1 7/1999
WO 03071410 A2 8/2003
(Continued)

OTHER PUBLICATIONS

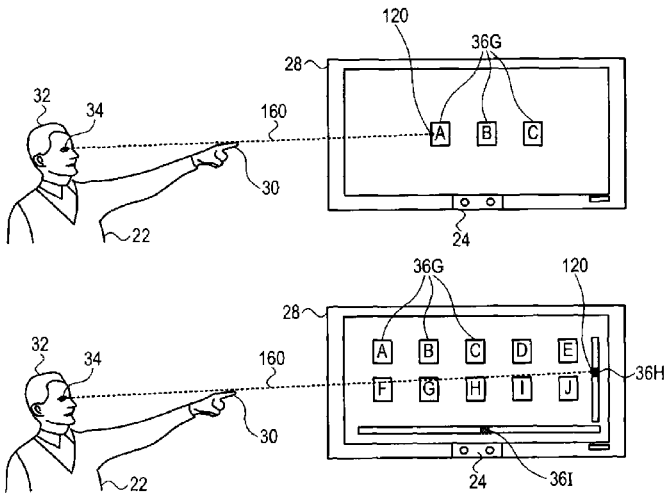
U.S. Appl. No. 13/849,514 Office Action dated Feb. 20, 2015.
(Continued)

Primary Examiner — Alexander Eisen
Assistant Examiner — Abhishek Sarma
(74) *Attorney, Agent, or Firm* — D. Kliger IP Services Ltd.

(57) **ABSTRACT**

A method includes receiving and segmenting a first sequence of three-dimensional (3D) maps over time of at least a part of a body of a user of a computerized system in order to extract 3D coordinates of a first point and a second point of the user, the 3D maps indicating a motion of the second point with respect to a display coupled to the computerized system. A line segment that intersects the first point and the second point is calculated, and a target point is identified where the line segment intersects the display. An interactive item presented on the display in proximity to the target point is engaged.

23 Claims, 10 Drawing Sheets



(56) **References Cited**

U.S. PATENT DOCUMENTS

4,988,981	A		1/1991	Zimmerman et al.
5,434,370	A		7/1995	Wilson et al.
5,454,043	A	*	9/1995	Freeman A61B 5/1121 345/419
5,495,576	A		2/1996	Ritchey
5,588,139	A		12/1996	Lanier et al.
5,594,469	A		1/1997	Freeman et al.
5,805,167	A		9/1998	Van Cruyningen
5,846,134	A		12/1998	Latypov
5,852,672	A		12/1998	Lu
5,862,256	A		1/1999	Zetts et al.
5,864,635	A		1/1999	Zetts et al.
5,870,196	A		2/1999	Lulli et al.
5,917,937	A		6/1999	Szeliski et al.
5,973,700	A		10/1999	Taylor et al.
6,002,808	A		12/1999	Freeman
6,084,979	A		7/2000	Kanade et al.
6,243,054	B1		6/2001	DeLuca
6,246,779	B1		6/2001	Fukui et al.
6,252,988	B1		6/2001	Ho
6,256,033	B1		7/2001	Nguyen
6,262,740	B1		7/2001	Lauer et al.
6,345,111	B1		2/2002	Yamaguchi et al.
6,345,893	B2		2/2002	Fateh et al.
6,452,584	B1		9/2002	Walker et al.
6,456,262	B1		9/2002	Bell
6,483,499	B1		11/2002	Li et al.
6,507,353	B1		1/2003	Huard et al.
6,512,838	B1		1/2003	Rafii et al.
6,519,363	B1		2/2003	Su et al.
6,559,813	B1		5/2003	DeLuca et al.
6,681,031	B2		1/2004	Cohen et al.
6,686,921	B1		2/2004	Rushmeier et al.
6,690,370	B2		2/2004	Ellenby et al.
6,741,251	B2		5/2004	Malzbender
6,791,540	B1		9/2004	Baumberg
6,803,928	B2		10/2004	Bimber et al.
6,853,935	B2		2/2005	Satoh et al.
6,857,746	B2		2/2005	Dyner
6,977,654	B2		12/2005	Malik et al.
7,003,134	B1		2/2006	Covell et al.
7,013,046	B2		3/2006	Kawamura et al.
7,023,436	B2		4/2006	Segawa et al.
7,042,440	B2		5/2006	Pryor et al.
7,042,442	B1		5/2006	Kanevsky et al.
7,151,530	B2		12/2006	Roeber et al.
7,170,492	B2		1/2007	Bell
7,215,815	B2		5/2007	Honda
7,227,526	B2		6/2007	Hildreth et al.
7,259,747	B2		8/2007	Bell
7,264,554	B2		9/2007	Bentley
7,289,227	B2		10/2007	Smetak et al.
7,301,648	B2		11/2007	Foxlin
7,302,099	B2		11/2007	Zhang et al.
7,333,113	B2		2/2008	Gordon
7,340,077	B2		3/2008	Gokturk
7,348,963	B2		3/2008	Bell
7,358,972	B2		4/2008	Gordon et al.
7,370,883	B2		5/2008	Basir et al.
7,427,996	B2		9/2008	Yonezawa et al.
7,428,542	B1		9/2008	Fink et al.
7,474,256	B2		1/2009	Ohta et al.
7,526,120	B2		4/2009	Gokturk et al.
7,536,032	B2		5/2009	Bell
7,573,480	B2		8/2009	Gordon
7,576,727	B2		8/2009	Bell
7,580,572	B2		8/2009	Bang et al.
7,590,941	B2		9/2009	Wee et al.
7,688,998	B2		3/2010	Tuma et al.
7,696,876	B2		4/2010	Dimmer et al.
7,724,250	B2		5/2010	Ishii et al.
7,762,665	B2		7/2010	Vertegaal et al.
7,774,155	B2		8/2010	Sato et al.
7,809,160	B2		10/2010	Vertegaal et al.
7,812,842	B2		10/2010	Gordon
7,821,541	B2		10/2010	Delean
7,840,031	B2		11/2010	Albertson et al.
7,844,914	B2		11/2010	Andre et al.
7,925,549	B2		4/2011	Looney et al.
7,971,156	B2		6/2011	Albertson et al.
8,150,142	B2		4/2012	Freedman et al.
8,166,421	B2		4/2012	Magal et al.
8,249,334	B2		8/2012	Berliner et al.
8,368,647	B2		2/2013	Lin
8,390,821	B2		3/2013	Shpunt et al.
8,400,494	B2		3/2013	Zalevsky et al.
8,493,496	B2		7/2013	Freedman et al.
8,514,221	B2		8/2013	King et al.
2002/0057383	A1		5/2002	Iwamura
2002/0071607	A1		6/2002	Kawamura et al.
2002/0158873	A1		10/2002	Williamson
2003/0057972	A1		3/2003	Pfaff et al.
2003/0063775	A1		4/2003	Rafii et al.
2003/0088463	A1		5/2003	Kanevsky
2003/0146901	A1		8/2003	Ryan
2003/0156756	A1		8/2003	Gokturk et al.
2003/0185444	A1		10/2003	Honda
2003/0234346	A1		12/2003	Kao
2003/0235341	A1		12/2003	Gokturk et al.
2004/0046744	A1		3/2004	Rafii et al.
2004/0104935	A1		6/2004	Williamson et al.
2004/0135744	A1		7/2004	Bimber et al.
2004/0165099	A1		8/2004	Stavely et al.
2004/0174770	A1		9/2004	Rees
2004/0183775	A1		9/2004	Bell
2004/0184640	A1		9/2004	Bang et al.
2004/0184659	A1		9/2004	Bang et al.
2004/0258314	A1		12/2004	Hashimoto
2005/0031166	A1		2/2005	Fujimura et al.
2005/0062684	A1		3/2005	Geng
2005/0088407	A1		4/2005	Bell et al.
2005/0089194	A1		4/2005	Bell
2005/0110964	A1		5/2005	Bell et al.
2005/0122308	A1		6/2005	Bell et al.
2005/0162381	A1		7/2005	Bell et al.
2005/0190972	A1		9/2005	Thomas et al.
2005/0254726	A1		11/2005	Fuchs et al.
2005/0265583	A1		12/2005	Covell et al.
2006/0010400	A1		1/2006	Dehlin et al.
2006/0092138	A1		5/2006	Kim et al.
2006/0110008	A1		5/2006	Vertegaal et al.
2006/0115155	A1		6/2006	Lui et al.
2006/0139314	A1		6/2006	Bell
2006/0149737	A1		7/2006	Du et al.
2006/0159344	A1		7/2006	Shao et al.
2006/0187196	A1		8/2006	Underkoffler et al.
2006/0239670	A1		10/2006	Cleveland
2006/0248475	A1		11/2006	Abrahamsson
2007/0078552	A1		4/2007	Rosenberg
2007/0154116	A1		7/2007	Shieh
2007/0230789	A1		10/2007	Chang et al.
2008/0019589	A1		1/2008	Yoon et al.
2008/0062123	A1		3/2008	Bell
2008/0094371	A1		4/2008	Forstall et al.
2008/0123940	A1		5/2008	Kundu et al.
2008/0150890	A1		6/2008	Bell et al.
2008/0150913	A1		6/2008	Bell et al.
2008/0170776	A1		7/2008	Albertson et al.
2008/0236902	A1		10/2008	Imaizumi
2008/0252596	A1		10/2008	Bell et al.
2008/0256494	A1		10/2008	Greenfield
2008/0260250	A1		10/2008	Vardi
2008/0273755	A1	*	11/2008	Hildreth G06F 1/1626 382/103
2008/0287189	A1		11/2008	Rabin
2009/0009593	A1		1/2009	Cameron et al.
2009/0027335	A1		1/2009	Ye
2009/0027337	A1		1/2009	Hildreth
2009/0031240	A1		1/2009	Hildreth
2009/0040215	A1		2/2009	Afzulpurkar et al.
2009/0077504	A1		3/2009	Bell
2009/0078473	A1		3/2009	Overgard et al.
2009/0083122	A1		3/2009	Angell et al.
2009/0083622	A1		3/2009	Chien et al.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2009/0096783	A1	4/2009	Shpunt et al.	
2009/0103780	A1*	4/2009	Nishihara et al.	G06F 3/017 382/103
2009/0183125	A1	7/2009	Magal et al.	
2009/0195392	A1	8/2009	Zalewski	
2009/0228841	A1	9/2009	Hildreth	
2009/0256817	A1	10/2009	Perlin et al.	
2009/0284608	A1	11/2009	Hong et al.	
2009/0297028	A1	12/2009	De Haan	
2010/0002936	A1	1/2010	Khomo et al.	
2010/0007717	A1	1/2010	Spektor et al.	
2010/0034457	A1	2/2010	Berliner et al.	
2010/0036717	A1	2/2010	Trest	
2010/0053151	A1	3/2010	Marti et al.	
2010/0071965	A1	3/2010	Hu et al.	
2010/0149096	A1	6/2010	Migos et al.	
2010/0162181	A1	6/2010	Shiplacoff et al.	
2010/0164897	A1	7/2010	Morin et al.	
2010/0177933	A1	7/2010	Willmann et al.	
2010/0194863	A1	8/2010	Lopes et al.	
2010/0199228	A1	8/2010	Latta et al.	
2010/0231504	A1	9/2010	Bloem et al.	
2010/0234094	A1	9/2010	Gagner et al.	
2010/0235786	A1	9/2010	Maizels et al.	
2011/0006978	A1	1/2011	Yuan	
2011/0018795	A1	1/2011	Jang	
2011/0029918	A1*	2/2011	Yoo et al.	715/800
2011/0052006	A1	3/2011	Gurman et al.	
2011/0081072	A1	4/2011	Kawasaki et al.	
2011/0164032	A1	7/2011	Shadmi	
2011/0164141	A1	7/2011	Tico et al.	
2011/0175932	A1	7/2011	Yu et al.	
2011/0178784	A1	7/2011	Sato et al.	
2011/0193939	A1	8/2011	Vassigh et al.	
2011/0211754	A1	9/2011	Litvak et al.	
2011/0225492	A1	9/2011	Boettcher et al.	
2011/0225536	A1	9/2011	Shams et al.	
2011/0227820	A1	9/2011	Haddick et al.	
2011/0231757	A1	9/2011	Haddick et al.	
2011/0248914	A1	10/2011	Sherr	
2011/0254765	A1	10/2011	Brand	
2011/0254798	A1	10/2011	Adamson et al.	
2011/0260965	A1	10/2011	Kim et al.	
2011/0261058	A1	10/2011	Luo	
2011/0279397	A1	11/2011	Rimon et al.	
2011/0291926	A1	12/2011	Gokturk et al.	
2011/0292036	A1	12/2011	Sali et al.	
2011/0293137	A1	12/2011	Gurman et al.	
2011/0310010	A1	12/2011	Hoffnung et al.	
2012/0001875	A1	1/2012	Li et al.	
2012/0019703	A1	1/2012	Thorn	
2012/0078614	A1	3/2012	Galor et al.	
2012/0113104	A1	5/2012	Jung et al.	
2012/0169583	A1	7/2012	Rippel et al.	
2012/0184335	A1	7/2012	Kim et al.	
2012/0202569	A1	8/2012	Maizels et al.	
2012/0204133	A1	8/2012	Guendelman et al.	
2012/0216151	A1*	8/2012	Sarkar et al.	G06F 3/04883 715/863
2012/0223882	A1	9/2012	Galor et al.	
2012/0313848	A1	12/2012	Galor et al.	
2013/0014052	A1	1/2013	Frey et al.	
2013/0044053	A1	2/2013	Galor et al.	
2013/0055120	A1	2/2013	Galor et al.	
2013/0055150	A1	2/2013	Galor	
2013/0058565	A1	3/2013	Rafii et al.	
2013/0106692	A1	5/2013	Maizels et al.	
2013/0107021	A1	5/2013	Maizels et al.	
2013/0127854	A1	5/2013	Shpunt et al.	
2013/0154913	A1*	6/2013	Genc et al.	345/156
2013/0155070	A1	6/2013	Luo	
2013/0201108	A1	8/2013	Hirsch	
2013/0207920	A1	8/2013	McCann et al.	
2013/0222239	A1	8/2013	Galor	
2013/0263036	A1	10/2013	Berenson et al.	
2013/0265222	A1	10/2013	Berenson et al.	
2013/0283208	A1	10/2013	Bychkov et al.	
2013/0283213	A1	10/2013	Guendelman et al.	

FOREIGN PATENT DOCUMENTS

WO	2004107272	A1	12/2004
WO	2005003948	A1	1/2005
WO	2005094958	A1	10/2005
WO	2007078639	A1	7/2007
WO	2007135376	A2	11/2007
WO	2009083984	A2	7/2009
WO	2012107892	A1	8/2012

OTHER PUBLICATIONS

Bohme et al., "Head tracking with combined face and nose detection", IEEE International Symposium on Signals, Circuits and Systems, IEEE, 4 pages, year 2009.

Breitenstein et al., "Real-time face pose estimation from single range images", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 8 pages, Jun. 2008.

Feldmann et al., "Imnnersive Multi-User 3d Video Communication", Proceedings of International Broadcast Conference (IBC), 8 pages, Sep. 2009.

Haker et al., "Geometric Invariants for Facial Feature Tracking with 3D TOF cameras", IEEE International Symposium on Signals, Circuits and Systems, vol. 1, 4 pages, year 2007.

Malassiotis et al., "Robust Real-Time 3D Head Pose Estimation from Range Date", Pattern Recognition 38.8, 31 pages, year 2005.

Stipes et al., "4D Scan Registration with the SR-3000 LIDAR", IEEE International Conference on Robotics and Automation, pp. 2988-2993, May 19-23, 2008.

Yip et al., "Pose Determination and Viewpoint Determination of Human Head in Video Conferencing Based on Head Movement", IEEE Proceedings of the 10th International Multimedia Modelling Conference, 6 pages, Jan. 2004.

Gordon et al., "The use of Dense Stereo Range Date in Augmented Reality", Proceedings of the 1st International Symposium on Mixed and Augmented Reality (ISMAR), Darmstadt, Germany, pp. 1-10, Sep. 30-Oct. 1, 2002.

Agrawala et al., "The two-user Responsive Workbench :support for collaboration through individual views of a shared space", Proceedings on the 24th conference on computer graphics and interactive techniques (SIGGRAPH 97), Los Angeles, USA, pp. 327-332 , Aug. 3-8, 1997.

Harman et al., "Rapid 2D-to 3D conversion", Proceedings of SPIE Conference on Stereoscopic Displays and Virtual Reality Systems, vol. 4660, pp. 78-86, Jan. 21-23, 2002.

Hoff et al., "Analysis of head pose accuracy in augmented reality", IEEE Transactions on Visualization and Computer Graphics, vol. 6, No. 4, pp. 319-334, Oct.-Dec. 2000.

Poupyrev et al., "The go-go interaction technique: non-liner mapping for direct manipulation in VR", Proceedings of the 9th annual ACM Symposium on User interface software and technology (UIST '96), Washington, USA, pp. 79-80, Nov. 6-8, 1996.

Wexelblat et al., "Virtual Reality Applications and Explorations", Academic Press Inc., San Diego, USA, 262 pages, year 1993.

U.S. Appl. No. 13/161,508 Office Action dated Apr. 10, 2013.

U.S. Appl. No. 12/683,452 Office Action dated Jun. 7, 2013.

Miller, R., "Kinect for XBox 360 Review", Engadget, Nov. 4, 2010.

U.S. Appl. No. 13/161,508 Office Action dated Sep. 9, 2013.

International Application PCT/IB2013/052332 Search Report dated Aug. 26, 2013.

U.S. Appl. No. 13/314,210 Office Action dated Jul. 19, 2013.

U.S. Appl. No. 13/314,207 Office Action dated Aug. 5, 2013.

Sun et al., "SRP Based Natural Interaction Between Real and Virtual Worlds in Augmented Reality", Proceedings of the International Conference on Cyberworlds (CW'08), pp. 117-124, Sep. 22-24, 2008.

Schmalstieg et al., "The Studierstube Augmented Reality Project", Presence: Teleoperators and Virtual Environments, vol. 11, No. 1, pp. 33-54, Feb. 2002.

(56) **References Cited**

OTHER PUBLICATIONS

Ohta et al., "Share-Z: Client/Server Depth Sensing for See-Through Head-Mounted Displays", Presence: Teleoperators and Virtual Environments, vol. 11, No. 2, pp. 176-188, Apr. 2002.
Gobbetti et al., "VB2: an Architecture for Interaction in Synthetic Worlds", Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology (UIST'93), pp. 167-178, Nov. 3-5, 1993.
Gargallo et al., "Bayesian 3D Modeling from Images Using Multiple Depth Maps", Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 2, pp. 885-891, Jun. 20-25, 2005.
ZRRO Ltd., "TeleTouch Device", year 2011 (http://www.zrro.com/products.html).
Berliner et al., U.S. Appl. No. 61/732,354, filed Dec. 12, 2012.
Shpunt et al., U.S. Appl. No. 61/764,554, filed Feb. 14, 2013.
Bleiwess et al., "Fusing Time-of-Flight Depth and Color for Real-Time Segmentation and Tracking", Dyn3D 2009, Lecture Notes in Computer Science 5742, pp. 58-69, Jena, Germany, Sep. 9, 2009.
Bleiwess et al., "Markerless Motion Capture Using a Single Depth Sensor", SIGGRAPH Asia 2009, Yokohama, Japan, Dec. 16-19, 2009.
Bevilacqua et al., "People Tracking Using a Time-of-Flight Depth Sensor", Proceedings of the IEEE International Conference on Video and Signal Based Surveillance, Sydney, Australia, Nov. 22-24, 2006.
Bradski, G., "Computer Vision Face Tracking for Use in a Perceptual User Interface", Intel Technology Journal, vol. 2, issue 2 (2nd Quarter 2008).
Comaniciu et al., "Kernel-Based Object Tracking", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, No. 5, pp. 564-577, May 2003.
Gesturetec Inc., "Gesture Control Solutions for Consumer Devices", Canada, 2009.
Gokturk et al., "A Time-Of-Flight Depth Sensor—System Description, Issues and Solutions", Proceedings of the 2004 Conference on Computer Vision and Patter Recognition Workshop (CVPRW'04), vol. 3, pp. 35, Jun. 27-Jul. 2, 2004.
Grest et al., "Single View Motion Tracking by Depth and Silhouette Information", SCIA 2007—Scandinavian Conference on Image Analysis, Lecture Notes in Computer Science 4522, pp. 719-729, Aalborg, Denmark, Jun. 10-14, 2007.
Haritaoglu et al., "Ghost 3d: Detecting Body Posture and Parts Using Stereo", Proceedings of the IEEE Workshop on Motion and Video Computing (Motion'02), pp. 175-180, Orlando, USA, Dec. 5-6, 2002.
Haritaoglu et al., "W4S : A real-time system for detecting and tracking people in 2<1/2>D", ECCV 98—5th European conference on computer vision, vol. 1407, pp. 877-892, Freiburg , Germany, Jun. 2-6, 1998.
Harville, M., "Stereo Person Tracking with Short and Long Term Plan-View Appearance Models of Shape and Color", Proceedings of the IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSSS-2005), pp. 522-527, Como, Italy, Sep. 15-16, 2005.
Holte, M., "Fusion of Range and Intensity Information for View Invariant Gesture Recognition", IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '08), pp. 1-7, Anchorage, USA, Jun. 23-28, 2008.
Kaewtrakulpong et al., "An Improved Adaptive Background Mixture Model for Real-Time Tracking with Shadow Detection", Proceedings of the 2nd European Workshop on Advanced Video Based Surveillance Systems (AVBS'01), Kingston, UK, Sep. 2001.
Kolb et al., "ToF-Sensors: New Dimensions for Realism and Interactivity", Proceedings of the IEEE Conference on Computer Vision and Patter Recognition Workshops, pp. 1-6, Anchorage, USA, Jun. 23-28, 2008.
Kolsch et al., "Fast 2D Hand Tracking with Flocks of Features and Multi-Cue Integration", IEEE Workshop on Real-Time Vision for Human Computer Interaction (at CVPR'04), Washington, USA, Jun. 27-Jul. 2, 2004.
Krumm et al., "Multi-Camera Multi-Person Tracking for EasyLiving", 3rd IEEE International Workshop on Visual Surveillance, Dublin, Ireland, Jul. 1, 2000.
Leens et al., "Combining Color, Depth, and Motion for Video Segmentation", ICVS 2009—7th International Conference on Computer Vision Systems, Liege, Belgium Oct. 13-15, 2009.
MacCormick et al., "Partitioned Sampling, Articulated Objects, and Interface-Quality Hand Tracking", ECCV'00—Proceedings of the 6th European Conference on Computer Vision-Part ll , pp. 3-19, Dublin, Ireland, Jun. 26-Jul. 1, 2000.
Malassiotis et al., "Real-Time Hand Posture Recognition Using Range Data", Image and Vision Computing, vol. 26, No. 7, pp. 1027-1037, Jul. 2, 2008.
Morano et al., "Structured Light Using Pseudorandom Codes", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, issue 3, pp. 322-327, Mar. 1998.
Munoz-Salinas et al., "People Detection and Tracking Using Stereo Vision and Color", Image and Vision Computing, vol. 25, No. 6, pp. 995-1007, Jun. 1, 2007.
Nanda et al., "Visual Tracking Using Depth Data", Proceedings of the 2004 Conference on Computer Vision and Patter Recognition Workshop, vol. 3, Washington, USA, Jun. 27-Jul. 2, 2004.
Scharstein et al., "High-Accuracy Stereo Depth Maps Using Structured Light", IEEE Conference on Computer Vision and Patter Recognition, vol. 1, pp. 195-2002, Madison, USA, Jun. 2003.
Shi et al., "Good Features to Track", IEEE Conference on Computer Vision and Pattern Recognition, pp. 593-600, Seattle, USA, Jun. 21-23, 1994.
Siddiqui et al., "Robust Real-Time Upper Body Limb Detection and Tracking", Proceedings of the 4th ACM International Workshop on Video Surveillance and Sensor Networks, Santa Barbara, USA, Oct. 27, 2006.
Softkinetic S.A., IISU™—3D Gesture Recognition Platform for Developers of 3D Applications, Belgium, Brussels, 2007-2010.
Sudderth et al., "Visual Hand Tracking Using Nonparametric Belief Propagation", IEEE Workshop on Generative Model Based Vision at CVPR'04, Washington, USA, Jun. 27-Jul. 2, 2004.
Tsap, L., "Gesture-Tracking in Real Time with Dynamic Regional Range Computation", Real-Time Imaging, vol. 8, issue 2, pp. 115-126, Apr. 2002.
Xu et al., "A Multi-Cue-Based Human Body Tracking System", Proceedings of the 5ths International Conference on Computer Vision Systems (ICVS 2007), Germany, Mar. 21-24, 2007.
Xu et al., "Human Detecting Using Depth and Gray Images", Proceedings of the IEE Conference on Advanced Video and Signal Based Surveillance (AVSS'03), Miami, USA, Jul. 21-22, 2003.
Yilmaz et al., "Object Tracking: A Survey", ACM Computing Surveys, vol. 38, No. 4, article 13, Dec. 2006.
Zhu et al., "Controlled Human Pose Estimation From Depth Image Streams", IEEE Conference on Computer Vision and Patter Recognition Workshops, pp. 1-8, Anchorage, USA, Jun. 23-27, 2008.
International Application PCT/IB2010/051055 Search Report dated Sep. 1, 2010.
La Viola, J. Jr., "Whole-Hand and Speech Input in Virtual Environments", Computer Science Department, Florida Atlantic University, USA, 1996.
Martell, C., "Form: An Experiment in the Annotation of the Kinematics of Gesture", Dissertation, Computer and Information Science, University of Pennsylvania, 2005.
U.S. Appl. No. 12/352,622 Official Action dated Mar. 31, 2011.
Prime Sense Inc., "Prime Sensor™ NITE 1.1 Framework Programmer's Guide", Version 1.2, year 2009.
Primesense Corporation, "PrimeSensor Reference Design 1.08", USA, year 2010.
International Application PCT/IB2011/053192 Search Report dated Dec. 6, 2011.
U.S. Appl. No. 12/352,622 Official Action dated Sep. 30, 2011.
Hart, D., U.S. Appl. No. 09/616,606 "Method and System for High Resolution , Ultra Fast 3-D Imaging" filed Jul. 14, 2000.
International Application PCT/IL2007/000306 Search Report dated Oct. 2, 2008.
International Application PCT/IL2007/000574 Search Report dated Sep. 10, 2008.
International Application PCT/IL2006/000335 Preliminary Report on Patentability dated Apr. 24, 2008.
Avidan et al., "Trajectory triangulation: 3D reconstruction of moving points from amonocular image sequence", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 22, No. 4, pp. 348-3537, Apr. 2000.

(56) **References Cited**

OTHER PUBLICATIONS

Leclerc et al., "The direct computation of height from shading", The Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 552-558, USA, Jun. 1991.
Zhang et al., "Shape from intensity gradient", IEEE Transactions on Systems, Man and Cybernetics—Part A: Systems and Humans, vol. 29, No. 3, pp. 318-325, May 1999.
Zhang et al., "Height recovery from intensity gradients", IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 508-513, Jun. 21-23, 1994.
Horn, B., "Height and gradient from shading", International Journal of Computer Vision, vol. 5, No. 1, pp. 37-76, Aug. 1990.
Bruckstein, A., "On shape from shading", Computer Vision, Graphics & Image Processing, vol. 44, pp. 139-154, year 1988.
Zhang et al., "Rapid Shape Acquisition Using Color Structured Light and Multi-Pass Dynamic Programming", 1st International Symposium on 3D Data Processing Visualization and Transmission (3DPVT), Italy, Jul. 2002.
Besl, P., "Active, Optical Range Imaging Sensors", Machine vision and applications, vol. 1, pp. 127-152, year 1988.
Horn et al., "Toward optimal structured light patterns", Proceedings of International Conference on Recent Advances in 3D Digital Imaging and Modeling, pp. 28-37, Ottawa, Canada, May 1997.
Goodman, J.W., "Statistical Properties of Laser Speckle Patterns", Laser Speckle and Related Phenomena, pp. 9-75, Springer-Verlag, Berlin Heidelberg, 1975.
Asada et al., "Determining Surface Orientation by Projecting a Stripe Pattern", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 10, No. 5, pp. 749-754, Sep. 1988.
Winkelbach et al., "Shape from Single Stripe Pattern Illumination", Luc Van Gool (Editor), (DAGM 2002), Lecture Notes in Computer Science 2449, p. 240-247, Springer 2002.
Koninckx et al., "Efficient, Active 3D Acquisition, based on a Pattern-Specific Snake", Luc Van Gool (Editor), (DAGM 2002), Lecture Notes in Computer Science 2449, pp. 557-565, Springer 2002.
Kimmel et al., "Analyzing and synthesizing images by evolving curves with the Osher-Sethian method", International Journal of Computer Vision, vol. 24, No. 1, pp. 37-56, year 1997.
Zigelman et al., "Texture mapping using surface flattening via multi-dimensional scaling", IEEE Transactions on Visualization and Computer Graphics, vol. 8, No. 2, pp. 198-207, Apr. 2002.
Dainty, J.C., "Introduction", Laser Speckle and Related Phenomena, pp. 1-7, Springer-Verlag, Berlin Heidelberg, 1975.
Mendlovic et al., "Composite harmonic filters for scale, projection and shift invariant pattern recognition", Applied Optics Journal, vol. 34, No. 2, Jan. 10, 1995.
Fua et al., "Human Shape and Motion Recovery Using Animation Models", 19th Congress, International Society for Photogrammetry and Remote Sensing, Amsterdam, The Netherlands, Jul. 2000.
Allard et al., "Marker-less Real Time 3D modeling for Virtual Reality", Immersive Projection Technology, Iowa State University, year 2004.
Howe et al., "Bayesian Reconstruction of 3D Human Motion from Single-Camera Video", Advanced in Neural Information Processing Systems, vol. 12, pp. 820-826, USA 1999.
Li et al., "Real-Time 3D Motion Tracking with Known Geometric Models", Real-Time Imaging Journal, vol. 5, pp. 167-187, Academic Press 1999.
Grammalidis et al., "3-D Human Body Tracking from Depth Images Using Analysis by Synthesis", Proceedings of the IEEE International Conference on Image Processing (ICIP2001), pp. 185-188, Greece, Oct. 7-10, 2001.
Segen et al., "Shadow gestures: 3D hand pose estimation using a single camera", Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 479-485, Fort Collins, USA, 1999.
Vogler et al., "ASL recognition based on a coupling between HMMs and 3D motion analysis", Proceedings of IEEE International Conference on Computer Vision, pp. 363-369, Mumbai, India, 1998.
Nam et al., "Recognition of Hand Gestures with 3D, Nonlinear Arm Movements", Pattern Recognition Letters, vol. 18, No. 1, pp. 105-113, Elsevier Science B.V. 1997.
Nesbat, S., "A System for Fast, Full-Text Entry for Small Electronic Devices", Proceedings of the 5th International Conference on Multimodal Interfaces, ICMI 2003, Vancouver, Nov. 5-7, 2003.
Ascension Technology Corporation, "Flock of Birds: Real-Time Motion Tracking", 2008.
Segen et al., "Human-computer interaction using gesture recognition and 3D hand tracking", ICIP 98, Proceedings of the IEEE International Conference on Image Processing, vol. 3, pp. 188-192, Oct. 4-7, 1998.
Dekker, L., "Building Symbolic Information for 3D Human Body Modeling from Range Data", Proceedings of the Second International Conference on 3D Digital Imaging and Modeling, IEEE computer Society, pp. 388-397, 1999.
Holte et al., "Gesture Recognition using a Range Camera", Technical Report CVMT-07-01 ISSN 1601-3646, Feb. 2007.
Cheng et al., "Articulated Human Body Pose Inference from Voxel Data Using a Kinematically Constrained Gaussian Mixture Model", CVPR EHuM2: 2nd Workshop on Evalutation of Articulated Human Motion and Pose Estimation, 2007.
Microvision Inc., "PicoP® Display Engine—How it Works", 1996-2012.
Primesense Corporation, "PrimeSensor NITE 1.1", USA, year 2010.
ARM Ltd., "AMBA Specification: AHB", Version 2, pp. 35-92, year 1999.
Commission Regulation (EC) No. 1275/2008, Official Journal of the European Union, Dec. 17, 2008.
Primesense, "Natural Interaction", YouTube Presentation, Jun. 9, 2010 http://www.youtube.com/watch?v=TzLKsex43zl~.
Manning et al., "Foundations of Statistical Natural Language Processing", chapters 6,7,9 and 12, MIT Press 1999.
U.S. Appl. No. 12/762,336 Official Action dated May 15, 2012.
Tobii Technology, "The World Leader in Eye Tracking and Gaze Interaction", Mar. 2012.
Noveron, "Madison video eyewear", year 2012.
International Application PCT/IB2012/050577 Search Report dated Aug. 6, 2012.
U.S. Appl. No. 12/683,452 Official Action dated Sep. 7, 2012.
Koutek, M., "Scientific Visualization in Virtual Reality: Interaction Techniques and Application Development", PhD Thesis, Delft University of Technology, 264 pages, Jan. 2003.
Azuma et al., "Recent Advances in Augmented Reality", IEEE Computer Graphics and Applications, vol. 21, issue 6, pp. 34-47, Nov. 2001.
Breen et al., "Interactive Occlusion and Collision of Real and Virtual Objects in Augmented Reality", Technical Report ECRC-95-02, ECRC, Munich, Germany, 22 pages, year 1995.
Burdea et al., "A Distributed Virtual Environment with Dextrous Force Feedback", Proceedings of Interface to Real and Virtual Worlds Conference, pp. 255-265, Mar. 1992.
U.S. Appl. No. 13/244,490 Office Action dated Dec. 6, 2013.
U.S. Appl. No. 13/423,314 Office Action dated Dec. 4, 2013.
U.S. Appl. No. 13/423,322 Office Action dated Nov. 1, 2013.
U.S. Appl. No. 13/984,031 Office Action dated Jul. 2, 2015.
U.S. Appl. No. 13/960,822 Office Action dated May 29, 2015.
CN Application # 201280007484.1 Office Action dated Jul. 8, 2015.
U.S. Appl. No. 13/849,517 Office Action dated Oct. 7, 2015.
Ridden, P., "Microsoft HoloDesk lets Users Handle Virtual 3D objects", 7 pages, Oct. 24, 2011 Published at http://www.gizmag.com/holodesk-lets-users-handle-virtual-3d-objects/20257/.
KR Application # 10-2014-7029627 Office Action dated Aug. 28, 2015.
U.S. Appl. No. 13/960,822 Office Action dated Aug. 13, 2015.
U.S. Appl. No. 13/849,514, Office Action dated Dec. 2, 2015.
U.S. Appl. No. 13/960,822, Office Action dated Nov. 24, 2015.
Irie et al., "Construction of an intelligent room based on gesture recognition: operation of electric appliances with hand gestures", Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 1, pp. 193-198, Sep. 28-Oct. 2, 2004.
CN Application # 201280007484.1 Office Action dated Dec. 15, 2015.

* cited by examiner

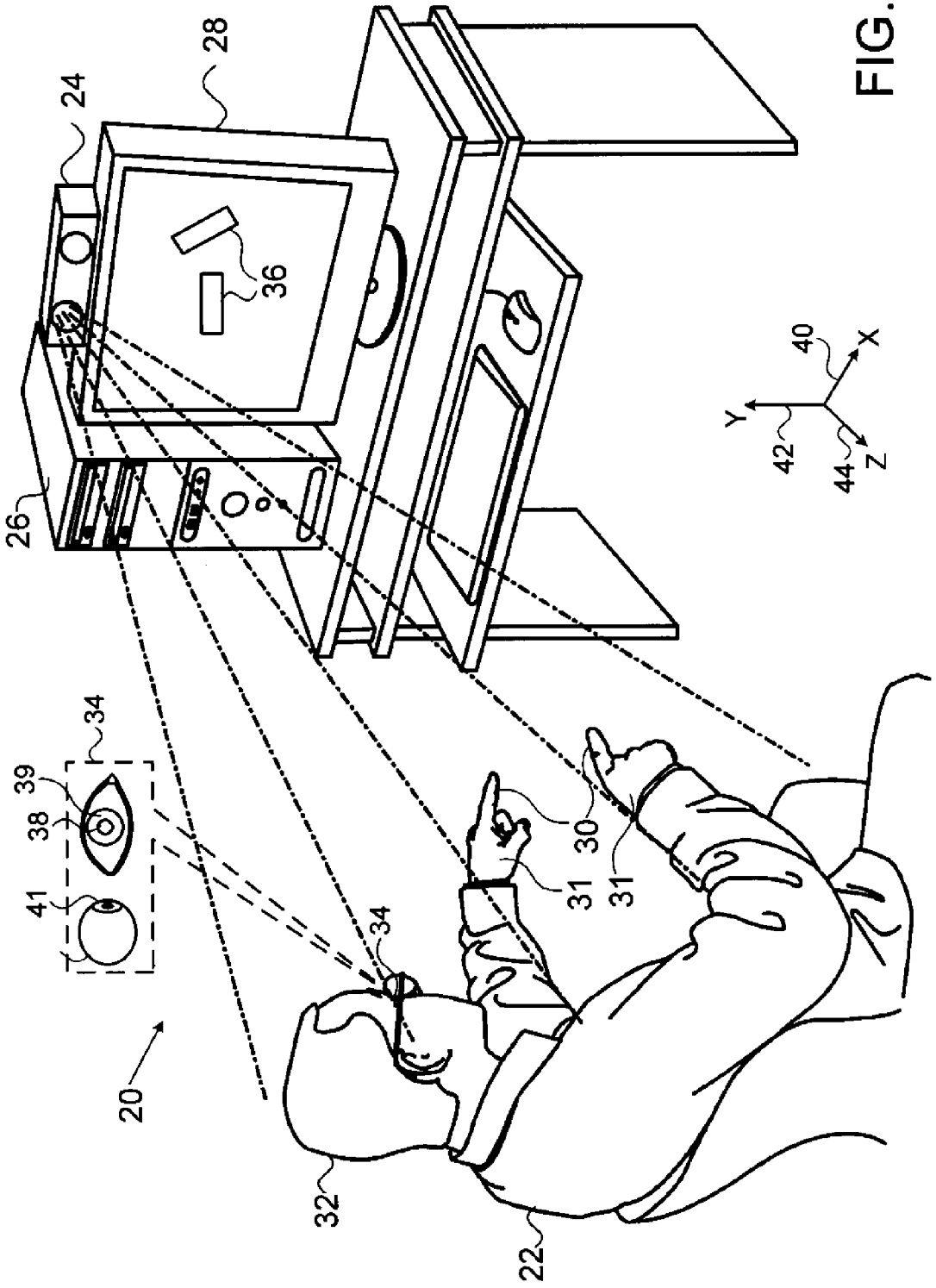


FIG. 1

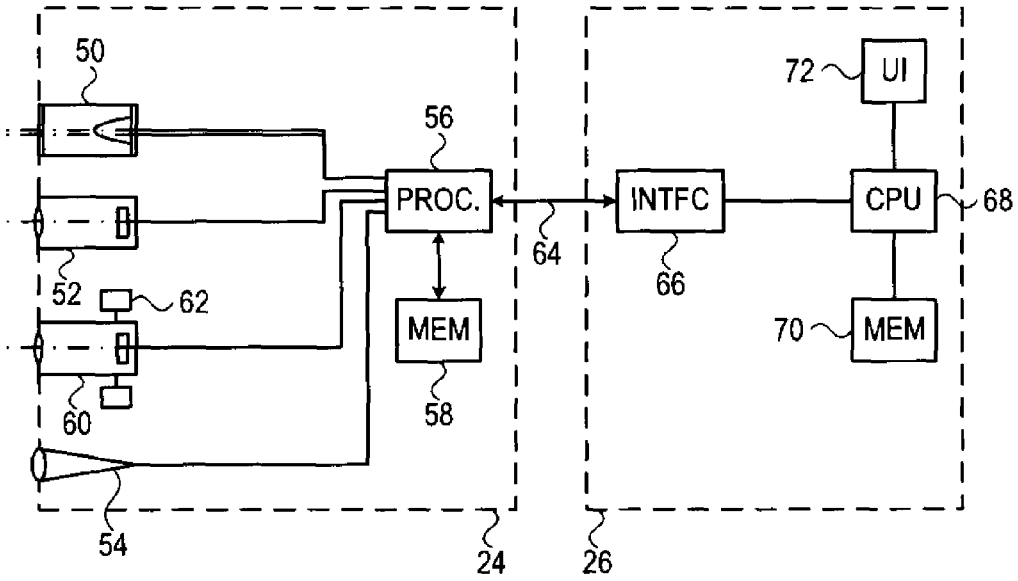

50
56
PROC.
64
INTFC
72
UI
CPU
68
52
62
MEM
66
70
MEM
60
58
54
24
26

FIG. 2

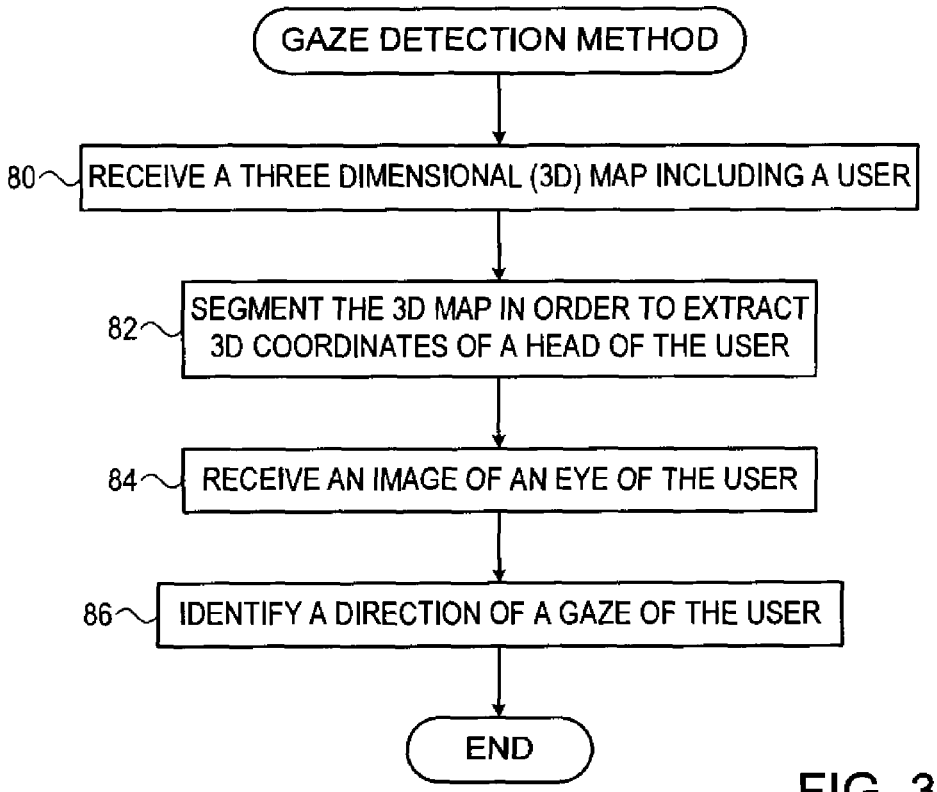

GAZE DETECTION METHOD
80
RECEIVE A THREE DIMENSIONAL (3D) MAP INCLUDING A USER
82
SEGMENT THE 3D MAP IN ORDER TO EXTRACT 3D COORDINATES OF A HEAD OF THE USER
84
RECEIVE AN IMAGE OF AN EYE OF THE USER
86
IDENTIFY A DIRECTION OF A GAZE OF THE USER
END

FIG. 3

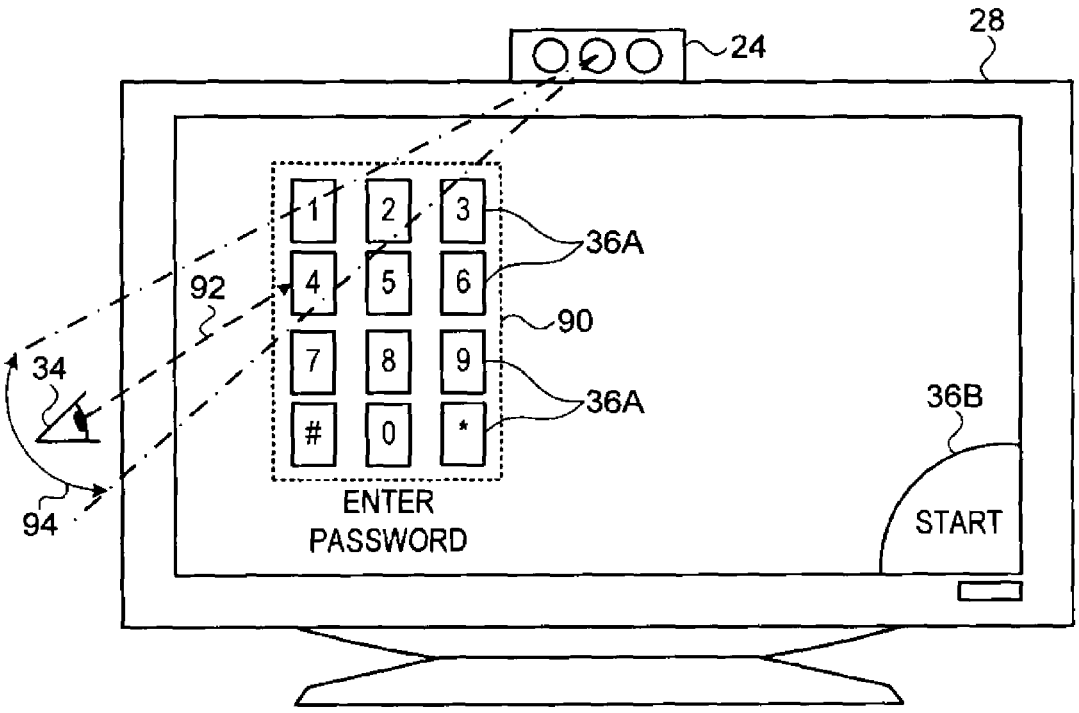


FIG. 4

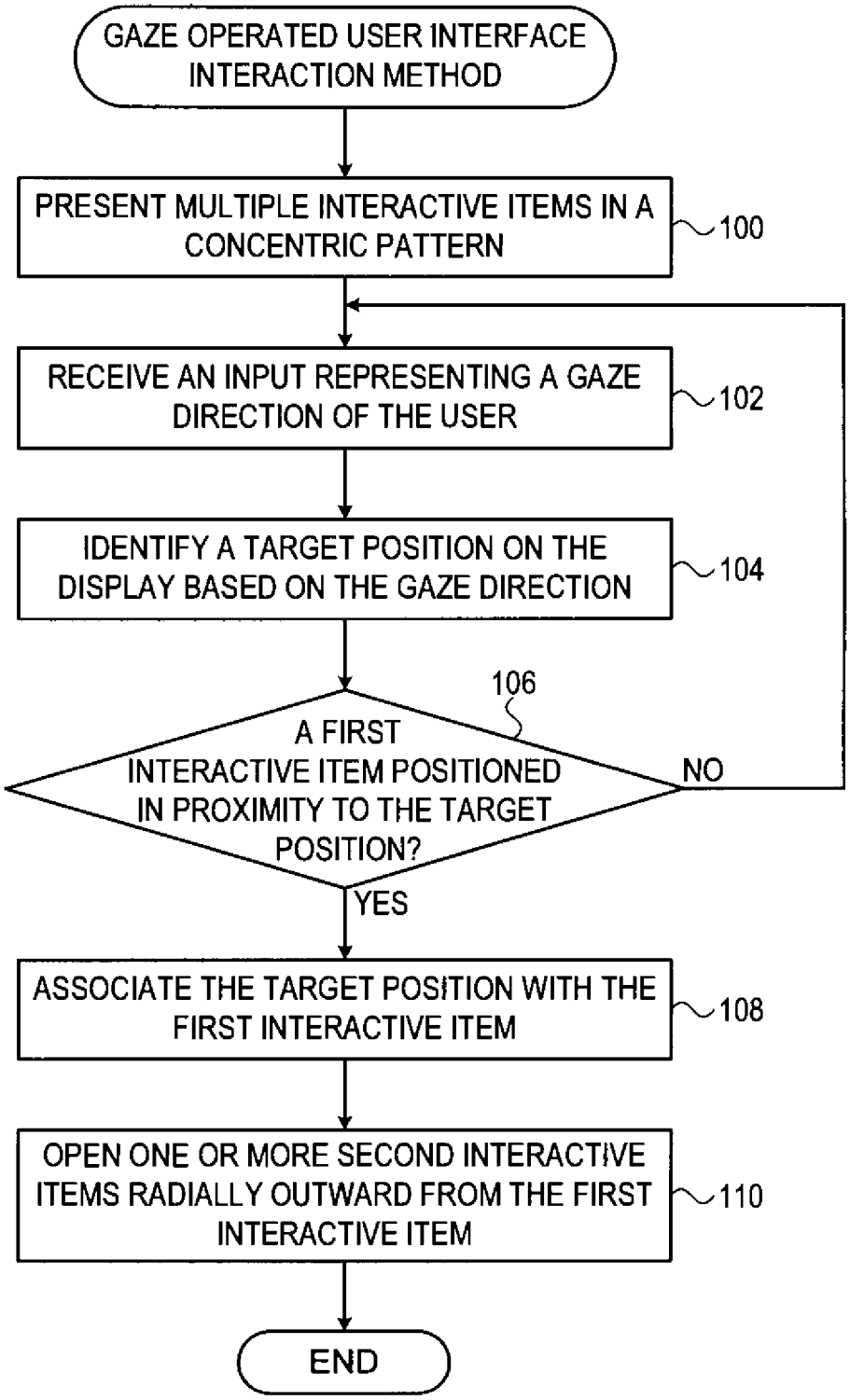

GAZE OPERATED USER INTERFACE INTERACTION METHOD
PRESENT MULTIPLE INTERACTIVE ITEMS IN A CONCENTRIC PATTERN
100
RECEIVE AN INPUT REPRESENTING A GAZE DIRECTION OF THE USER
102
IDENTIFY A TARGET POSITION ON THE DISPLAY BASED ON THE GAZE DIRECTION
104
106
A FIRST INTERACTIVE ITEM POSITIONED IN PROXIMITY TO THE TARGET POSITION?
NO
YES
ASSOCIATE THE TARGET POSITION WITH THE FIRST INTERACTIVE ITEM
108
OPEN ONE OR MORE SECOND INTERACTIVE ITEMS RADIALLY OUTWARD FROM THE FIRST INTERACTIVE ITEM
110
END

FIG. 5

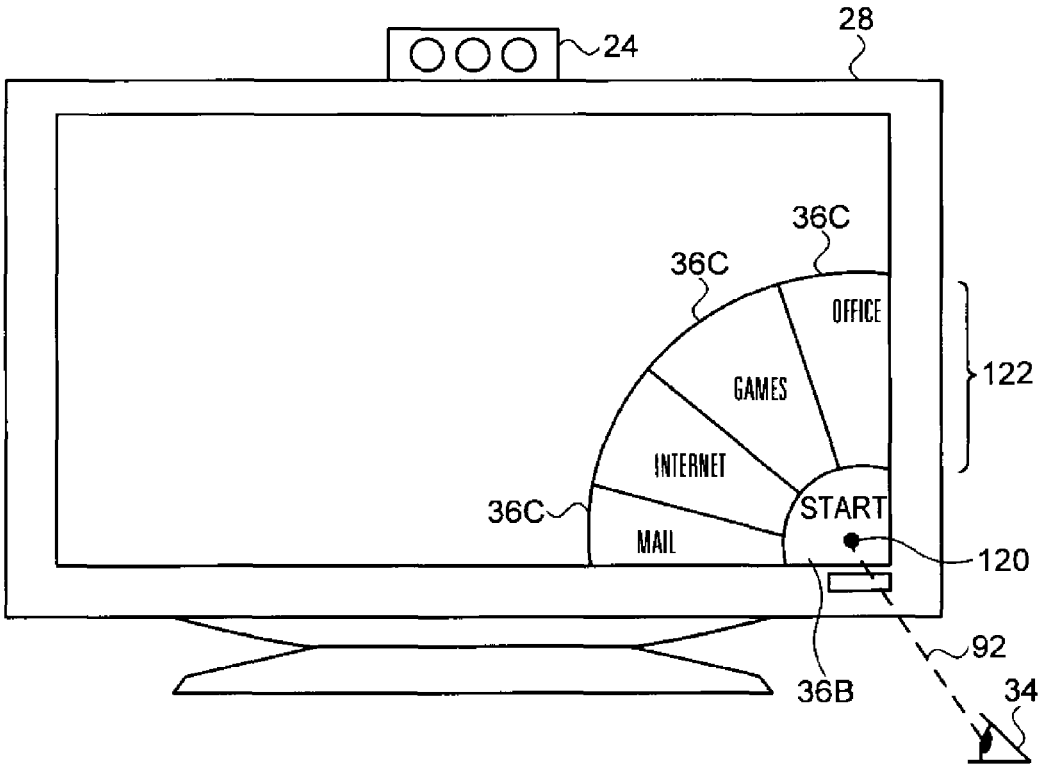

24
28
36C
36C
36C
OFFICE
GAMES
INTERNET
MAIL
START
122
120
92
34
36B

FIG. 6A

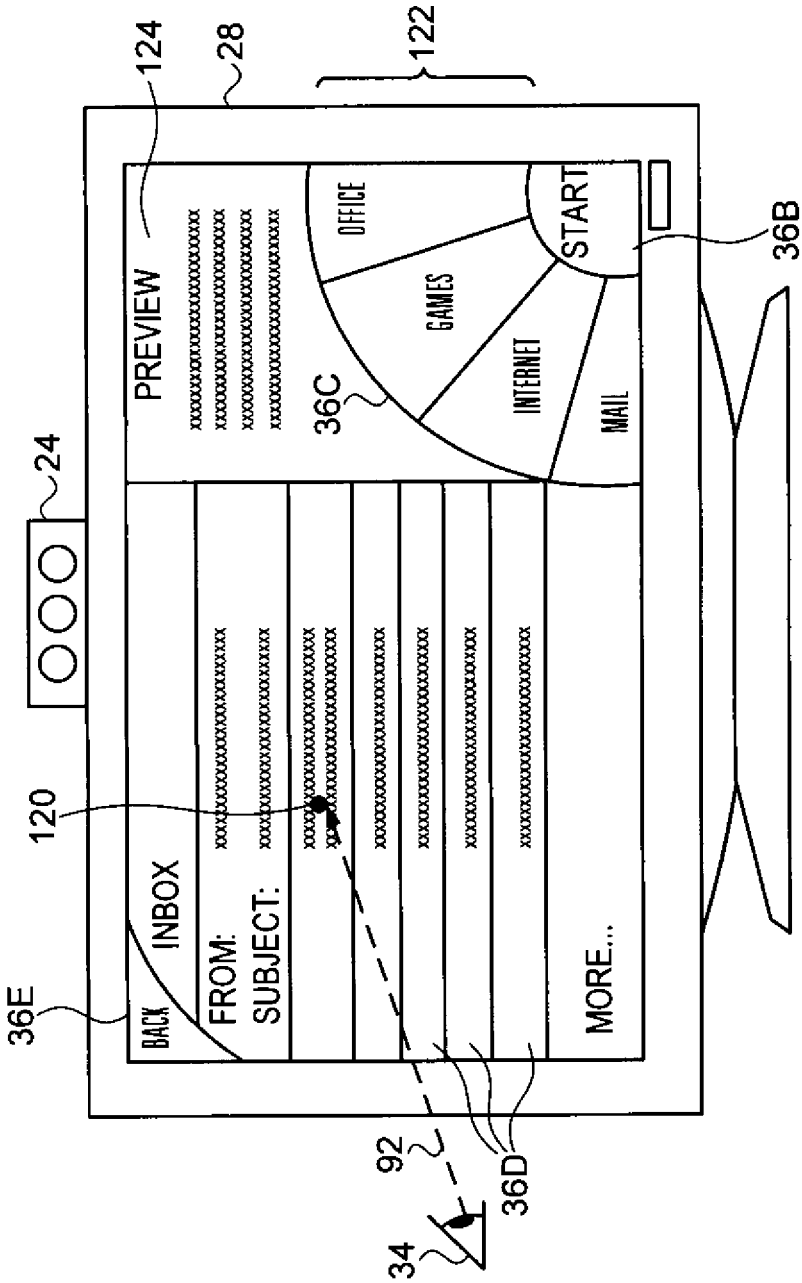

124
28
122
PREVIEW
OFFICE
START
36B
GAMES
INTERNET
MAIL
36C
24
120
INBOX
FROM:
SUBJECT:
BACK
36E
MORE...
92
36D
34

FIG. 6B

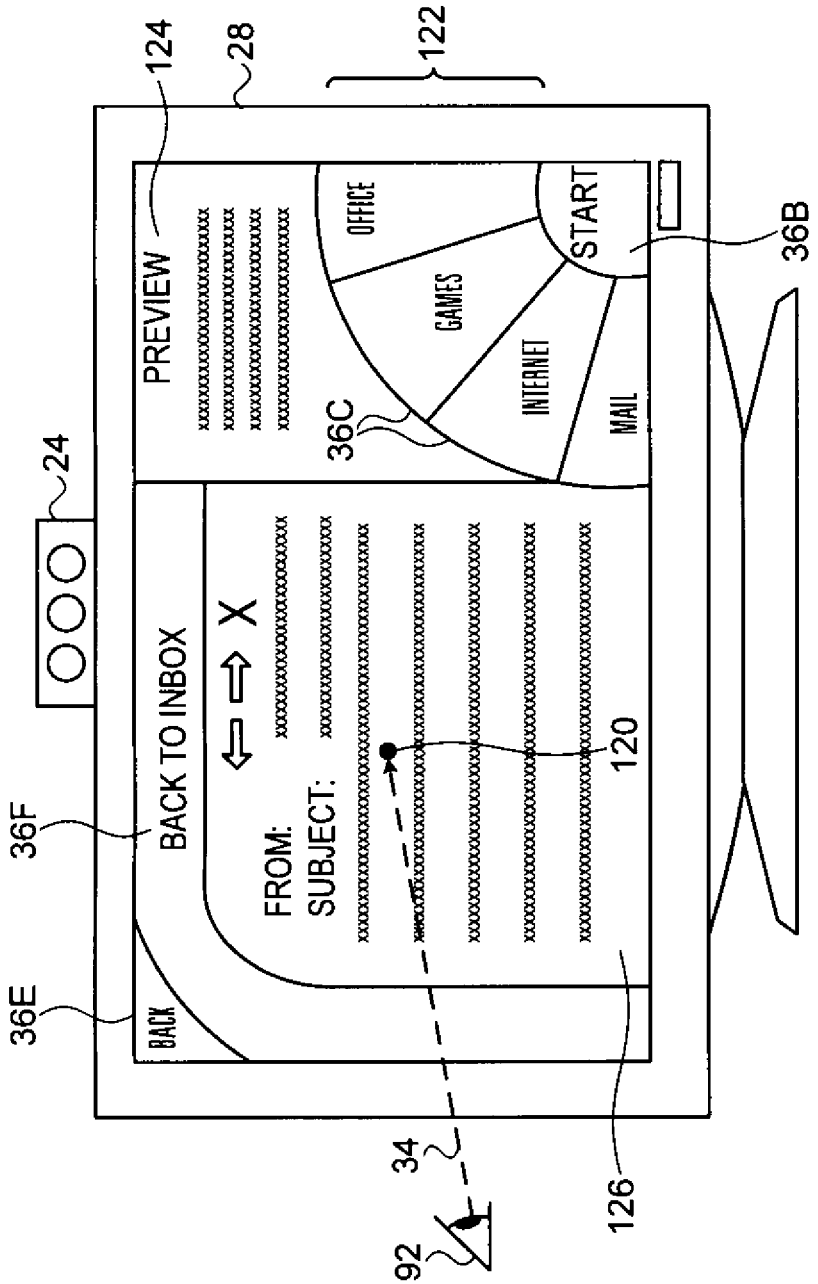
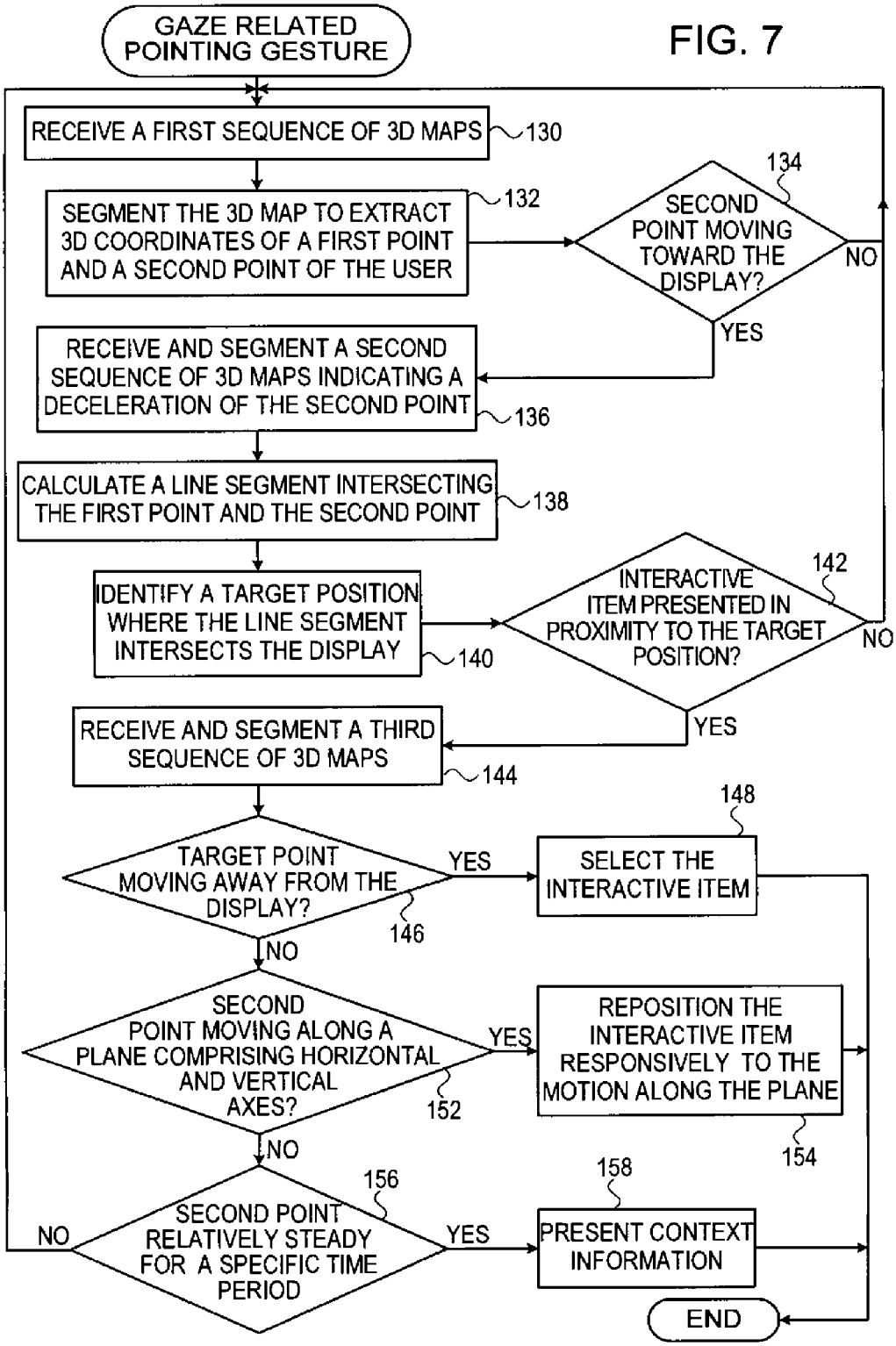

124
28
122
PREVIEW
OFFICE
GAMES
INTERNET
MAIL
START
36C
36B
24
36F
BACK TO INBOX
FROM:
SUBJECT:
120
36E
BACK
34
92
126

FIG. 6C

FIG. 7


GAZE RELATED POINTING GESTURE
RECEIVE A FIRST SEQUENCE OF 3D MAPS
130
SEGMENT THE 3D MAP TO EXTRACT 3D COORDINATES OF A FIRST POINT AND A SECOND POINT OF THE USER
132
SECOND POINT MOVING TOWARD THE DISPLAY?
134
NO
YES
RECEIVE AND SEGMENT A SECOND SEQUENCE OF 3D MAPS INDICATING A DECELERATION OF THE SECOND POINT
136
CALCULATE A LINE SEGMENT INTERSECTING THE FIRST POINT AND THE SECOND POINT
138
IDENTIFY A TARGET POSITION WHERE THE LINE SEGMENT INTERSECTS THE DISPLAY
140
INTERACTIVE ITEM PRESENTED IN PROXIMITY TO THE TARGET POSITION?
142
NO
YES
RECEIVE AND SEGMENT A THIRD SEQUENCE OF 3D MAPS
144
TARGET POINT MOVING AWAY FROM THE DISPLAY?
146
YES
148
SELECT THE INTERACTIVE ITEM
NO
SECOND POINT MOVING ALONG A PLANE COMPRISING HORIZONTAL AND VERTICAL AXES?
152
YES
REPOSITION THE INTERACTIVE ITEM RESPONSIVELY TO THE MOTION ALONG THE PLANE
154
NO
SECOND POINT RELATIVELY STEADY FOR A SPECIFIC TIME PERIOD
156
NO
YES
158
PRESENT CONTEXT INFORMATION
END

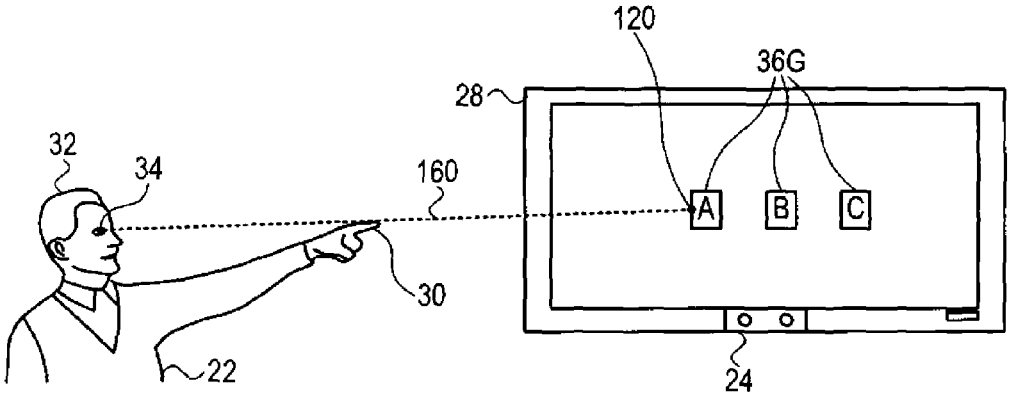


FIG. 8

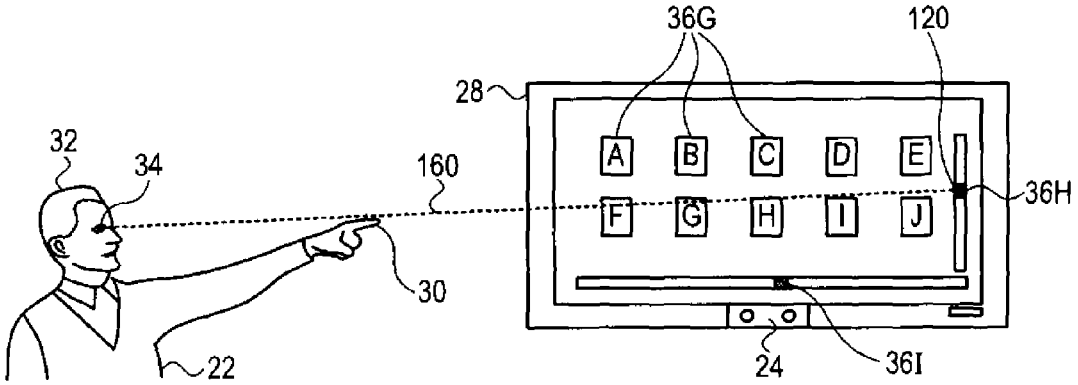


FIG. 9

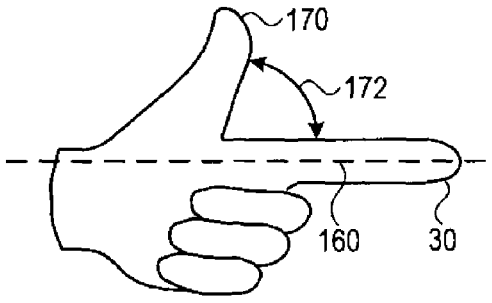


FIG. 10

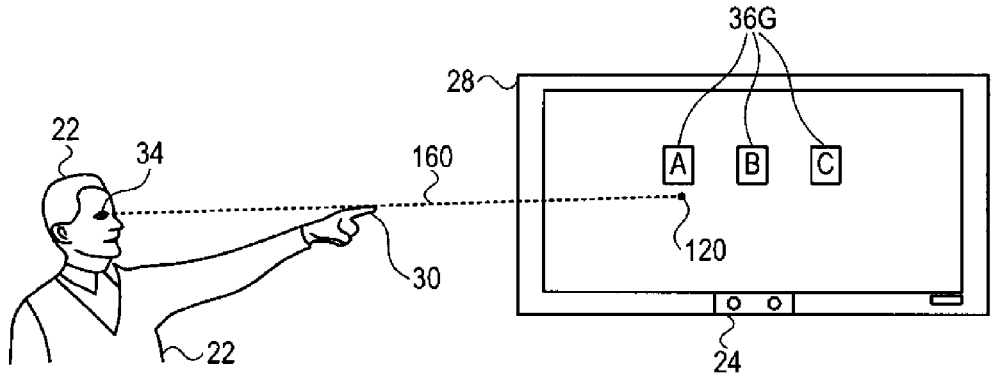


FIG. 11

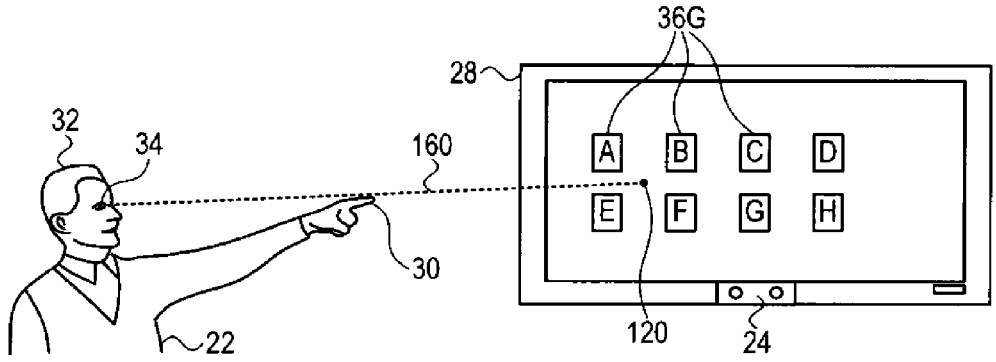


FIG. 12A

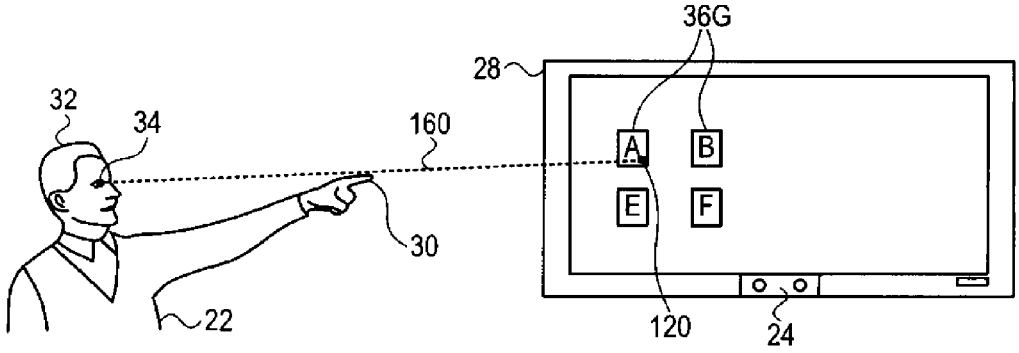


FIG. 12B

POINTING-BASED DISPLAY INTERACTION

RELATED APPLICATIONS

This application is a continuation of PCT Patent Application PCT/IB2012/050577, filed Feb. 9, 2012, which claims the benefit of U.S. Provisional Patent Application 61/440,877, filed on Feb. 9, 2011, U.S. Provisional Patent Application 61/526,692, filed on Aug. 24, 2011, and U.S. Provisional Patent Application 61/538,867, filed on Sep. 25, 2011, which are incorporated herein by reference.

FIELD OF THE INVENTION

The present invention relates generally to human-machine interfaces, and specifically to interfaces that combine multiple user interaction modalities.

BACKGROUND

Many different types of user interface devices and methods are currently available. Common tactile interface devices include a computer keyboard, a mouse and a joystick. Touch screens detect the presence and location of a touch by a finger or other object within the display area. Infrared remote controls are widely used, and "wearable" hardware devices have been developed, as well, for purposes of remote control.

Computer interfaces based on three-dimensional (3D) sensing of parts of a user's body have also been proposed. For example, PCT International Publication WO 03/071410, whose disclosure is incorporated herein by reference, describes a gesture recognition system using depth-perceptive sensors. A 3D sensor, typically positioned in a room in proximity to the user, provides position information, which is used to identify gestures created by a body part of interest. The gestures are recognized based on the shape of the body part and its position and orientation over an interval. The gesture is classified for determining an input into a related electronic device.

Documents incorporated by reference in the present patent application are to be considered an integral part of the application except that to the extent any terms are defined in these incorporated documents in a manner that conflicts with the definitions made explicitly or implicitly in the present specification, only the definitions in the present specification should be considered.

As another example, U.S. Pat. No. 7,348,963, whose disclosure is incorporated herein by reference, describes an interactive video display system, in which a display screen displays a visual image, and a camera captures 3D information regarding an object in an interactive area located in front of the display screen. A computer system directs the display screen to change the visual image in response to changes in the object.

Three-dimensional human interface systems may identify not only the user's hands, but also other parts of the body, including the head, torso and limbs. For example, U.S. Patent Application Publication 2010/0034457, whose disclosure is incorporated herein by reference, describes a method for modeling humanoid forms from depth maps. The depth map is segmented so as to find a contour of the body. The contour is processed in order to identify a torso and one or more limbs of the subject. An input is generated to control an application program running on a computer by analyzing a disposition of at least one of the identified limbs in the depth map.

Some user interface systems track the direction of the user's gaze. For example, U.S. Pat. No. 7,762,665, whose disclosure is incorporated herein by reference, describes a method of modulating operation of a device, comprising: providing an attentive user interface for obtaining information about an attentive state of a user; and modulating operation of a device on the basis of the obtained information, wherein the operation that is modulated is initiated by the device. Preferably, the information about the user's attentive state is eye contact of the user with the device that is sensed by the attentive user interface.

SUMMARY

There is provided, in accordance with an embodiment of the present invention a method including receiving a three-dimensional (3D) map of at least a part of a body of a user of a computerized system, receiving a two dimensional (2D) image of the user, the image including an eye of the user, extracting, from the 3D map and the 2D image, 3D coordinates of a head of the user, and identifying, based on the 3D coordinates of the head and the image of the eye, a direction of a gaze performed by the user.

There is also provided, in accordance with an embodiment of the present invention a method including receiving an image including an eye of a user of a computerized system, identifying, based the image of the eye, a direction of a gaze performed by the user, identifying, based on the direction of the gaze, a region on a display coupled to the computerized system, and performing an operation on content presented in the region.

There is additionally provided, in accordance with an embodiment of the present invention a method including presenting, by a computerized system, multiple interactive items on a display coupled to the computer, receiving, from a sensing device coupled to the computer, an input representing a gaze direction of a user, identifying a target point on the display based on the gaze direction, associating the target point with a first interactive item appearing on the display, and responsively to the target point, opening one or more second interactive items on the display.

There is further provided, in accordance with an embodiment of the present invention a method including receiving and segmenting a first sequence of three-dimensional (3D) maps over time of at least a part of a body of a user of a computerized system in order to extract 3D coordinates of a first point and a second point of the user, the 3D maps indicating a motion of the second point with respect to a display coupled to the computerized system, calculating a line segment that intersects the first point and the second point, identifying a target point where the line segment intersects the display, and engaging an interactive item presented on the display in proximity to the target point.

There is also provided, in accordance with an embodiment of the present invention an apparatus including a sensing device configured to receive a three dimensional (3D) map of at least a part of a body of a user and an image of an eye of the user, and to receive a two dimensional (2D) image of the user, the 2D image including an eye of the user, and a computer coupled to the sensing device and configured to extract, from the 3D map and the 2D image, 3D coordinates of a head of the user and to identify, based on the 3D coordinates of the head and the image of the eye, a direction of a gaze performed by the user.

There is additionally provided, in accordance with an embodiment of the present invention an apparatus including a sensing device configured to receive an image including an eye of a user, and a computer configured to identify, based the image of the eye, a direction of a gaze performed by the user,

to identify, based on the direction of the gaze, a region on a display coupled to the computerized system, and to perform an operation on content presented in the region.

There is further provided, in accordance with an embodiment of the present invention an apparatus including a display, and a computer coupled to the display and configured to present multiple interactive items on the display, to receive, from a sensing device coupled to the computer, an input representing a gaze direction of a user, to identify a target point on the display based on the gaze direction, to associate the target point with a first interactive item appearing on the display, and responsively to the target point, to open one or more second interactive items on the display.

There is also provided, in accordance with an embodiment of the present invention an apparatus including a display, and a computer coupled to the display and configured to receive and segment a first sequence of three-dimensional (3D) maps over time of at least a part of a body of a user of a computerized system in order to extract 3D coordinates of a first point and a second point of the user, the 3D maps indicating a motion of the second point with respect to a display coupled to the computer system, to calculate a line segment that intersects the first point and the second point, to identify a target point where the line segment intersects the display, and to engage an interactive item presented on the display in proximity to the target point.

There is additionally provided, in accordance with an embodiment of the present invention a computer software product including a non-transitory computer-readable medium, in which program instructions are stored, which instructions, when read by a computer, cause the computer to receive a three-dimensional (3D) map of at least a part of a body of a user of the computer, to receive a two dimensional (2D) image of the user, the image including an eye of the user, to extract, from the 3D map and the 2D image, 3D coordinates of a head of the user, and to identify, based on the 3D coordinates of the head and the image of the eye, a direction of a gaze performed by the user.

There is further provided, in accordance with an embodiment of the present invention a computer software product including a non-transitory computer-readable medium, in which program instructions are stored, which instructions, when read by a computer, cause the computer to receive an image including an eye of a user of the computer system, to identify, based the image of the eye, a direction of a gaze performed by the user, to identify, based on the direction of the gaze, a region on a display coupled to the computerized system, and to perform an operation on content presented in the region.

There is also provided, in accordance with an embodiment of the present invention a computer software product including a non-transitory computer-readable medium, in which program instructions are stored, which instructions, when read by a computer, cause the computer to present multiple interactive items on a display coupled to the computer, to receive, from a sensing device coupled to the computer, an input representing a gaze direction of a user, to identify a target point on the display based on the gaze direction, to associate the target point with a first interactive item appearing on the display, and responsively to the target point, to open one or more second interactive items on the display.

There is additionally provided, in accordance with an embodiment of the present invention a computer software product including a non-transitory computer-readable medium, in which program instructions are stored, which instructions, when read by a computer, cause the computer to receive and segment a first sequence of three-dimensional (3D) maps over time of at least a part of a body of a user of the computer in order to extract 3D coordinates of a first point and a second point of the user, the 3D maps indicating a motion of the second point with respect to a display coupled to the computer, to calculate a line segment that intersects the first point and the second point, to identify a target point where the line segment intersects the display, and to engage an interactive item presented on the display in proximity to the target point.

BRIEF DESCRIPTION OF THE DRAWINGS

The disclosure is herein described, by way of example only, with reference to the accompanying drawings, wherein:

FIG. **1** is a schematic, pictorial illustration of a computer system implementing a mixed-modality user interface, in accordance with an embodiment of the present invention;

FIG. **2** is a block diagram that schematically illustrates functional components of the computer system implementing the mixed-modality user interface, in accordance with an embodiment of the present invention;

FIG. **3** is a flow diagram that schematically illustrates a method of detecting a direction of a gaze of a user, in accordance with an embodiment of the present invention;

FIG. **4** is a schematic representation of a numeric keypad configured for entering a password, in accordance with an embodiment of the present invention;

FIG. **5** is a flow diagram that schematically illustrates a method of interacting with a gaze operated user interface, in accordance with an embodiment of the present invention;

FIGS. **6**A-**6**C are schematic representations illustrating a sequence of operations performed using the gaze operated user interface, in accordance with an embodiment of the present invention;

FIG. **7** is a flow diagram that schematically illustrates a method of detecting a gaze related pointing gesture, in accordance with an embodiment of the present invention;

FIG. **8** is a schematic pictorial illustration of the user performing a Point-Select gesture to select a first given interactive item, in accordance with an embodiment of the present invention;

FIG. **9** is a schematic pictorial illustration of the user performing a Point-Touch gesture to manipulate a second given interactive item, in accordance with an embodiment of the present invention;

FIG. **10** is a schematic pictorial illustration showing an alternative Point-Select gesture, also referred to herein as a Trigger gesture, in accordance with an embodiment of the present invention;

FIG. **11** is a schematic pictorial illustration of the user pointing a hand at a given icon presented on a display, in order to calibrate the computer system, in accordance with an embodiment of the present invention; and

FIGS. **12**A and **12**B are schematic pictorial illustrations of the computer system assisting the user to select a given icon by presenting the icons in smaller and larger sizes, in accordance with an embodiment of the present invention.

DETAILED DESCRIPTION OF EMBODIMENTS

Overview

When using physical tactile input devices such as buttons, rollers or touch screens, a user typically engages and disengages control of a user interface by touching and/or manipulating the physical device. Embodiments of the present invention describe pointing gestures for engaging interactive items

presented on a display coupled to a computer executing a mixed modality user interface that includes three-dimensional (3D) sensing, by a 3D sensor, of motion or change of position of one or more body parts, typically a hand or a finger, of the user. Pointing gestures described herein include Point-Select, Point-Touch and Point-Hold gestures that are explained in detail hereinbelow.

The Point-Select gesture enables the user to select an interactive item presented on the display. For example, using the Point-Select gesture, the user can start watching a movie by performing the Point-Select gesture toward an icon (on the display) associated with the move. The Point-Touch gesture enables the user to manipulate an interactive item presented on the display. For example, the user can horizontally scroll a list of interactive items (e.g., movies) presented on the display by manipulating a horizontal scroll box via the Point-Touch gesture. The Point-Hold gesture enables the user to view context information for an interactive item presented on the display. For example, in response to the user performing a Point-Hold gesture on an icon representing a movie, the computer can present a pop-up window including information such as a plot summary, a review, and cast members. In some embodiments, the mixed modality user interface may also convey visual feedback as the user performs the pointing gestures described supra.

While interacting with traditional two-dimensional (2D) user interfaces, the physical devices described supra typically convey tactile feedback to the user. However, while interacting with a 3D user interface such as the mixed modality user interface described herein, the user may perform gestures without engaging any physical device, and therefore not receive any tactile feedback. Embodiments of the present invention provide methods and systems for interacting with items presented on a display, and receiving non-tactile feedback, thereby compensating for the lack of tactile feedback.

System Description

FIG. **1** is a schematic, pictorial illustration of a mixed-modality user interface **20** for operation by a user **22** of a computer **26**, in accordance with an embodiment of the present invention. (Although for the sake of simplicity, only a single user and user interface are shown in the figure, in practice interface **20** may interact with multiple users concurrently. Alternative embodiments of the present invention may use different user interfaces and/or support multiple user interfaces across different devices). User interface **20** in the pictured embodiment is based, by way of example, on a 3D sensing device **24**, which captures 3D scene information that includes a body, or at least parts of the body, such as a finger **30**, a hand **31**, a head **32**, or eyes **34**. Device **24** or a separate camera (not shown in the figures) may also capture color video images of the scene. The information captured by device **24** is processed by computer **26**, which drives a display screen **28** accordingly to present and manipulate on-screen interactive items **36** (also referred to herein as interactive items). Alternatively, the user interface may be used in conjunction with any type of computerized equipment, such as a laptop, a tablet computer, a television, etc.

While FIG. **1** shows computer **26** in a tower configuration, other configurations of the computer are considered to be within the spirit and scope of the present invention. For example, computer **26** may be configured as a desktop computer, a portable computer (e.g., a laptop) or an all-in-one computer.

Computer **26** processes data generated by device **24** in order to reconstruct a 3D map of user **22**. The term "3D map" (or equivalently, "depth map") refers to a set of 3D coordinates representing a surface of a given object, in this case the user's body. In one embodiment, device **24** projects a pattern of spots onto the object and captures an image of the projected pattern. Computer **26** then computes the 3D coordinates of points on the surface of the user's body by triangulation, based on transverse shifts of the spots in the imaged pattern. The 3D coordinates are measured, by way of example, with reference to a generally horizontal X-axis **40**, a generally vertical Y-axis **42** and a depth Z-axis **44**, based on device **24**. Methods and devices for this sort of triangulation-based 3D mapping using a projected pattern are described, for example, in PCT International Publications WO 2007/043036, WO 2007/105205 and WO 2008/120217, whose disclosures are incorporated herein by reference. Alternatively, system **20** may use other methods of 3D mapping, using single or multiple cameras or other types of sensors, as are known in the art.

In some embodiments, device **24** detects the location and direction of eyes **34** of user **22**, typically by processing and analyzing an image comprising light (typically infrared and/or a color produced by the red-green-blue additive color model) reflecting from one or both eyes **34**, in order to find a direction of the user's gaze. In alternative embodiments, computer **26** (either by itself or in combination with device **24**) detects the location and direction of the eyes **34** of the user. The reflected light may originate from a light projecting source of device **24**, or any other natural (e.g., sunlight) or artificial (e.g., a lamp) source. Using techniques that are known in the art such as detecting pupil center and corneal reflections (PCCR), device **24** may process and analyze an image comprising light reflecting from an element of eye **34**, such as a pupil **38**, an iris **39** or a cornea **41**, in order to find the direction of the user's gaze. Additionally, device **24** may convey (to computer **26**) the light reflecting from the cornea as a glint effect.

The location and features of the user's head (e.g., an edge of the eye, a nose or a nostril) that are extracted by computer **26** from the 3D map may be used in finding coarse location coordinates of the user's eyes, thus simplifying the determination of precise eye position and gaze direction, and making the gaze measurement more reliable and robust. Furthermore, computer **26** can readily combine the 3D location of parts of head **32** (e.g., eye **34**) that are provided by the 3D map with gaze angle information obtained via eye part image analysis in order to identify a given on-screen object **36** at which the user is looking at any given time. This use of 3D mapping in conjunction with gaze tracking allows user **22** to move head **32** freely while alleviating the need to actively track the head using sensors or emitters on the head, as in some eye tracking systems that are known in the art.

By tracking eye **34**, embodiments of the present invention may reduce the need to re-calibrate user **22** after the user moves head **32**. In some embodiments, computer **26** may use depth information for head **32**, eye **34** and pupil **38**, in order to track the head's movement, thereby enabling a reliable gaze angle to be calculated based on a single calibration of user **22**. Utilizing techniques that are known in the art such as PCCR, pupil tracking, and pupil shape, computer **26** may calculate a gaze angle of eye **34** from a fixed point of head **32**, and use the head's location information in order to re-calculate the gaze angle and enhance the accuracy of the aforementioned techniques. In addition to reduced recalibrations, further benefits of tracking the head may include reducing the number of light projecting sources and reducing the number of cameras used to track eye **34**.

Computer **26** typically comprises a general-purpose computer processor, which is programmed in software to carry out

the functions described hereinbelow. The software may be downloaded to the processor in electronic form, over a network, for example, or it may alternatively be provided on non-transitory tangible computer-readable media, such as optical, magnetic, or electronic memory media. Alternatively or additionally, some or all of the functions of the computer processor may be implemented in dedicated hardware, such as a custom or semi-custom integrated circuit or a programmable digital signal processor (DSP). Although computer **26** is shown in FIG. **1**, by way of example, as a separate unit from sensing device **24**, some or all of the processing functions of the computer may be performed by suitable dedicated circuitry within the housing of the sensing device or otherwise associated with the sensing device.

As another alternative, these processing functions may be carried out by a suitable processor that is integrated with display **28** (in a television set, for example) or with any other suitable sort of computerized device, such as a game console or a media player. The sensing functions of device **24** may likewise be integrated into the computer or other computerized apparatus that is to be controlled by the sensor output.

Various techniques may be used to reconstruct the 3D map of the body of user **22**. In one embodiment, computer **26** extracts 3D connected components corresponding to the parts of the body from the depth data generated by device **24**. Techniques that may be used for this purpose are described, for example, in U.S. patent application Ser. No. 12/854,187, filed Aug. 11, 2010, whose disclosure is incorporated herein by reference. The computer analyzes these extracted components in order to reconstruct a "skeleton" of the user's body, as described in the above-mentioned U.S. Patent Application Publication 2010/0034457, or in U.S. patent application Ser. No. 12/854,188, filed Aug. 11, 2010, whose disclosure is also incorporated herein by reference. In alternative embodiments, other techniques may be used to identify certain parts of the user's body, and there is no need for the entire body to be visible to device **24** or for the skeleton to be reconstructed, in whole or even in part.

Using the reconstructed skeleton, computer **26** can assume a position of a body part such as a tip of finger **30**, even though the body part (e.g., the fingertip) may not be detected by the depth map due to issues such as minimal object size and reduced resolution at greater distances from device **24**. In some embodiments, computer **26** can auto-complete a body part based on an expected shape of the human part from an earlier detection of the body part, or from tracking the body part along several (previously) received depth maps.

In some embodiments, the information generated by computer **26** as a result of this skeleton reconstruction includes the location and direction of the user's head, as well as of the arms, torso, and possibly legs, hands and other features, as well. Changes in these features from frame to frame (i.e. depth maps) or in postures of the user can provide an indication of gestures and other motions made by the user. User posture, gestures and other motions may provide a control input for user interaction with interface **20**. These body motions may be combined with other interaction modalities that are sensed by device **24**, including user eye movements, as described above, as well as voice commands and other sounds. Interface **20** thus enables user **22** to perform various remote control functions and to interact with applications, interfaces, video programs, images, games and other multimedia content appearing on display **28**.

FIG. **2** is a block diagram that schematically illustrates functional components of user interface **20**, in accordance with an embodiment of the present invention. Sensing device **24** comprises an illumination subassembly **50**, which projects a pattern onto the scene of interest. A depth imaging subassembly **52**, such as a suitably-configured video camera, captures images of the pattern on the scene. Typically, illumination subassembly **50** and imaging subassembly **52** operate in the infrared range, although other spectral ranges may also be used. Optionally, a color video camera (not shown) in device **24** captures 2D color images of the scene, and a microphone **54** may also capture sound.

A processor **56** receives the images from subassembly **52** and compares the pattern in each image to a reference pattern stored in a memory **58**. The reference pattern is typically captured in advance by projecting the pattern onto a reference plane at a known distance from device **24**. Processor **56** computes local shifts of parts of the pattern over the area of the 3D map and translates these shifts into depth coordinates. Details of this process are described, for example, in PCT International Publication WO 2010/004542, whose disclosure is incorporated herein by reference. Alternatively, as noted earlier, device **24** may be configured to generate 3D maps by other means that are known in the art, such as stereoscopic imaging, sonar-like devices (sound based/acoustic), wearable implements, lasers, or time-of-flight measurements.

Processor **56** typically comprises an embedded microprocessor, which is programmed in software (or firmware) to carry out the processing functions that are described hereinbelow. The software may be provided to the processor in electronic form, over a network, for example; alternatively or additionally, the software may be stored on non-transitory tangible computer-readable media, such as optical, magnetic, or electronic memory media. Processor **56** also comprises suitable input and output interfaces and may comprise dedicated and/or programmable hardware logic circuits for carrying out some or all of its functions. Details of some of these processing functions and circuits that may be used to carry them out are presented in the above-mentioned Publication WO 2010/004542.

In some embodiments, a gaze sensor **60** detects the gaze direction of eyes **34** of user **22** by capturing and processing two dimensional images of user **22**. In alternative embodiments, computer **26** detects the gaze direction by processing a sequence of 3D maps conveyed by device **24**. Sensor **60** may use any suitable method of eye tracking that is known in the art, such as the method described in the above-mentioned U.S. Pat. No. 7,762,665 or in U.S. Pat. No. 7,809,160, whose disclosure is incorporated herein by reference, or the alternative methods described in references cited in these patents. For example, sensor **60** may capture an image of light (typically infrared light) that is reflected from the fundus and/or the cornea of the user's eye or eyes. This light may be projected toward the eyes by illumination subassembly **50** or by another projection element (not shown) that is associated with sensor **60**. Sensor **60** may capture its image with high resolution over the entire region of interest of user interface **20** and may then locate the reflections from the eye within this region of interest. Alternatively, imaging subassembly **52** may capture the reflections from the user's eyes (ambient light, reflection from monitor) in addition to capturing the pattern images for 3D mapping.

As another alternative, processor **56** may drive a scan control **62** to direct the field of view of gaze sensor **60** toward the location of the user's face or eye **34**. This location may be determined by processor **60** or by computer **26** on the basis of a depth map or on the basis of the skeleton reconstructed from the 3D map, as described above, or using methods of image-based face recognition that are known in the art. Scan control **62** may comprise, for example, an electromechanical gimbal,

or a scanning optical or optoelectronic element, or any other suitable type of scanner that is known in the art, such as a microelectromechanical system (MEMS) based mirror that is configured to reflect the scene to gaze sensor **60**.

In some embodiments, scan control **62** may also comprise an optical or electronic zoom, which adjusts the magnification of sensor **60** depending on the distance from device **24** to the user's head, as provided by the 3D map. The above techniques, implemented by scan control **62**, enable a gaze sensor **60** of only moderate resolution to capture images of the user's eyes with high precision, and thus give precise gaze direction information.

In alternative embodiments, computer **26** may calculate the gaze angle using an angle (i.e., relative to Z-axis **44**) of the scan control. In additional embodiments, computer **26** may compare scenery captured by the gaze sensor **60**, and scenery identified in 3D depth maps. In further embodiments, computer **26** may compare scenery captured by the gaze sensor **60** with scenery captured by a 2D camera having a wide field of view that includes the entire scene of interest. Additionally or alternatively, scan control **62** may comprise sensors (typically either optical or electrical) configured to verify an angle of the eye movement.

Processor **56** processes the images captured by gaze sensor **60** in order to extract the user's gaze angle. By combining the angular measurements made by sensor **60** with the 3D location of the user's head provided by depth imaging subassembly **52**, the processor is able to derive accurately the user's true line of sight in 3D space. The combination of 3D mapping with gaze direction sensing reduces or eliminates the need for precise calibration and comparing multiple reflection signals in order to extract the true gaze direction. The line-of-sight information extracted by processor **56** enables computer **26** to identify reliably the interactive item at which the user is looking.

The combination of the two modalities can allow gaze detection without using an active projecting device (i.e., illumination subassembly **50**) since there is no need for detecting a glint point (as used, for example, in the PCCR method). Using the combination can solve the glasses reflection known problem of other gaze methods that are known in the art. Using information derived from natural light reflection, the 2D image (i.e. to detect the pupil position), and the 3D depth map (i.e., to identify the head's position by detecting the head's features), computer **26** can calculate the gaze angle and identify a given interactive item **36** at which the user is looking.

As noted earlier, gaze sensor **60** and processor **56** may track either one or both of the user's eyes. If both eyes **34** are tracked with sufficient accuracy, the processor may be able to provide an individual gaze angle measurement for each of the eyes. When the eyes are looking at a distant object, the gaze angles of both eyes will be parallel; but for nearby objects, the gaze angles will typically converge on a point in proximity to an object of interest. This point may be used, together with depth information, in extracting 3D coordinates of the point on which the user's gaze is fixed at any given moment.

As mentioned above, device **24** may create 3D maps of multiple users who are in its field of view at the same time. Gaze sensor **60** may similarly find the gaze direction of each of these users, either by providing a single high-resolution image of the entire field of view, or by scanning of scan control **62** to the location of the head of each user.

Processor **56** outputs the 3D maps and gaze information via a communication link **64**, such as a Universal Serial Bus (USB) connection, to a suitable interface **66** of computer **26**. The computer comprises a central processing unit (CPU) **68** with a memory **70** and a user interface **72**, which drives display **28** and may include other components, as well. As noted above, device **24** may alternatively output only raw images, and the 3D map and gaze computations described above may be performed in software by CPU **68**. Middleware for extracting higher-level information from the 3D maps and gaze information may run on processor **56**, CPU **68**, or both. CPU **68** runs one or more application programs, which drive user interface **72** based on information provided by the middleware, typically via an application program interface (API). Such applications may include, for example, games, entertainment, Web surfing, and/or office applications.

Although processor **56** and CPU **68** are shown in FIG. **2** as separate functional elements with a certain division of processing tasks between them, the functions of the processor and CPU may alternatively be carried out by a single processing unit, or these functions may be divided among three or more processing units. Furthermore, although device **24** is shown as containing a certain combination of components in a particular arrangement, other device configurations may be used for the purposes described herein, and are considered to be within the scope of the present invention.

Gaze Detection

FIG. **3** is a flow diagram that schematically illustrates a method of detecting a direction of a gaze of user **22**, in accordance with an embodiment of the present invention. In an initial step **80**, computer **26** receives, from depth imaging subassembly **52**, a 3D map including at least a body part of user **22**, and in an extraction step **82**, the computer segments the received 3D map in order to extract 3D coordinates of head **32**. In a receive step **84**, computer **26** receives, from gaze sensor **60**, a two dimensional image of the user including eye **34**. As described supra, illumination subassembly **50** may project a light toward user **22**, and the received image may comprise light reflected off the fundus and/or the cornea of eye(s) **34**. In an identification step **86**, computer **26** analyzes the received depth map and image in order to identify a gaze direction of user **22**, and the method ends.

In some embodiments, computer **26** extracts the 3D coordinates of head **32** by identifying, from the 3D map, a position of the head along X-axis **40**, Y-axis **42** and Z-axis **44**. In alternative embodiments, computer **26** extracts the 3D coordinates of head **32** by identifying, from the 2D image a first position of the head along X-axis **40** and Y-axis **42**, and identifying, from the 3D map, a second position of the head along Z-axis **44**.

Embodiments of the present invention can use gaze detection to controlling a function of a computerized system such as computer **26** responsively to the direction of the gaze. In some embodiments, computer **26** can identify a given interactive item **36** presented, on display **28**, at a position in the direction of the gaze, and change a state of the given interactive item responsively to the direction of the gaze.

In a first embodiment, changing the state of the given interactive item **36** may comprise performing an operation associated with the given interactive item. For example, interactive items **36** may comprise menu choices that user **22** can select to present specific content (e.g., a movie or a television show) on display **28**. In a second embodiment, computer **26** can change the state of a given interactive item **36** by directing input received from the user to the given interactive item. For example, the given interactive item may comprise a text box, and if the user is gazing at the text box, computer **26** can direct any alphanumeric input received from a keyboard to the text box.

In alternative embodiments, computer **26** can identify a given interactive item **36** presented, on display **28**, at a position in the direction of the gaze, and change a state of the given interactive item responsively to a vocal command received from the user. For example, the given interactive item may comprise an icon associated with a software application, and the user can gaze at the icon and say the word "start" to execute the application. In additional embodiments, user interface **20** may be configured to identify a given interactive item **36** responsively to the direction of the gaze, and to manipulate the given interactive item responsively to a gesture performed by a limb (e.g., finger **30** or hand **31**). For example, after selecting the given interactive item, if the computer receives a sequence of 3D maps indicating that the user is moving hand **30** in a swiping motion (i.e., along a plane comprising X-axis **40** and Y-axis **42**), computer **26** can responsively reposition the selected interactive item in the direction of the swipe (e.g., left to right).

In further embodiments, embodiments of the present invention may be used to receive an image (either a 2D image or a 3D map) including eye **34**, identify, based on the image, a gaze direction, identify, based on the gaze direction a region on the display and in the direction of the gaze, and perform an operation on the region. For example, computer **26** may comprise a tablet computer incorporating a digital camera, display **28** may comprise a display for the tablet, and the camera may be configured to focus the camera's lens on an item presented in the identified region.

Additionally or alternatively, computer **26** can identify a device coupled to the computer and positioned in the direction of the gaze, and controlling a function of the device responsively to the direction of the gaze. For example, if the user gazes at a top of a speaker coupled to the computer, the computer can raise the volume level of the speaker, and if the user gazes at the bottom of the speaker, the computer can lower the volume level.

FIG. **4** is a schematic representation of a numeric keypad **90** presented on display **28**, configured for entering a password, in accordance with an embodiment of the present invention. The presence and gaze of user **22** in this and subsequent figures is represented by eye **34**, having a line of sight **92**. Sensing device **24** determines the line of sight by finding the 3D location of the user's head and the gaze direction of eye **34** within a field of view **94** of the device, as described above.

In the description herein, interactive items **36** may be differentiated by using a different description (e.g., icon or scroll box instead of interactive item) and appending a letter to the identifying numeral. For example, in FIG. **4**, interactive items **46** comprise numeric input keys **36**A and a start button **36**B. Upon presenting the input keys and the start button, computer **26** prompts user **22** to enter a password. Although numeric keys are shown in FIG. **4**, any sort of on-screen graphical elements may be used as "keys" for this purpose. When the user wishes to initiate operation of computer **26** or of a particular application executing on the computer, the user can direct his gaze from one given key **36**A to the next in the proper sequence in order to enter a password. Optionally, the user may make a particular hand gesture or voice command (such as "ENTER") as his gaze rests on each given key **36**A (i.e., in a sequence of keys comprising the password), in order to inform device **24** to record the selection.

In some embodiments, computer **26** can be configured to select a sequence of interactive items **36** (e.g., as the user moves his gaze from one interactive item to the next) with or without any additional input (i.e., a gesture or sound) made by the user to indicate which of the interactive item has been selected. Similarly, computer **26** can be configured not to provide any visual or audio feedback to the user indicating the selection. As a result, even if the user enters a password in a public place, onlookers will be unable to determine or copy the password.

After entering the password, the user may direct his gaze toward start button **36**B in order to continue interacting with the computer.

In addition to password entry, gaze detection can be used to enhance security in other ways. For example, computer **26** may learn the user's characteristic eye movement patterns and/or other biometric features of the eyes **34** as an additional means of identification. As another example, device **24** may be configured to find the gaze angle not only of user **22**, but also of other people positioned within field of view **94** (as identified by skeleton reconstruction from the 3D maps, for instance). In this case, device **24** may be configured to alert user **22** when another person is looking at display **28** (and may even prompt computer **26** to display an image of this other person, captured by device **24**, on the display). This sort of functionality can assist user **22** in protecting himself from eavesdropping and in deterring eavesdroppers from continuing to look at the display.

As a power saving feature, device **24** may detect when the user is not looking at display **28**, and computer **26** may activate power saving techniques when the user looks away for more than some threshold period of time. For example, computer **26** may dim or darken the display (i.e., decrease the brightness) entirely when the user is not looking at the display. When the user looks back toward the display, the computer may deactivate the power saving techniques. For example, computer **26** may increase the brightness to return it to full brightness upon detecting that user **22** returned his gaze toward display. This sort of gaze-dependent screen control is also useful in enhancing battery life of portable devices and reducing power consumption generally for cost saving and environmental friendliness.

Gaze Operated User Interface

Gaze tracking may be used to create an interactive user interface that can detect which on-screen interactive item the user is looking at (such as a text box, or an application such as a word processor), thereby obviating the need for a mouse and/or a keyboard. For example, when the user types text, the text is automatically directed to the text box at which the user is looking. As another example, when the user makes the first keystroke in a word or sentence, a "mouse click" type event is sent to the looked-at text box, which causes the text to be typed into the text box. In this manner the user may fill in a Web form without the need to move the mouse to switch from field to field ("first name" to "last name" to "password," etc.).

Furthermore, the combination of gaze tracking with other modalities, such as 3D mapping/gesture detection and/or voice detection, enables the user to control on-screen objects fully, without the use of a mouse or a touch screen. In this manner, the user can perform a full range of pointing and selection functions, including searching through large numbers of information items and choices. The combined interface modalities may also be used to search and perform control functions within the context of a certain interactive item, such as performing find, cut, copy and paste functions within an open file.

FIG. **5** is a flow diagram that schematically illustrates a method of interacting with a gaze operated user interface, in accordance with an embodiment of the present invention, and FIGS. **6**A-**6**C are schematic representations illustrating a

sequence of operations performed using a gaze operated user interface, in accordance with an embodiment of the present invention. In an initialization step **100**, a computerized system such as computer **26** presents multiple interactive items **36** on display **28**. In the configuration shown in FIGS. **6**A-**6**C, computer **26** initially presents start button **36**B in the lower right corner of display **28**. As user **22** directs his line of sight **92** toward start button **36**B, computer **26** receives an input (e.g., depth maps) indicating a gaze direction of user **22** in a receive step **102**. Upon determining the gaze direction, computer **26** identifies a target point **120** on display **28** based on the gaze direction (i.e., a point on the display that the user is looking at) in an identification step **104**.

In a comparison step **106**, if a first interactive item **36** is in proximity to target point, then in an association step **108**, computer **26** associates the target point with the first interactive item. In the example shown in FIG. **6**A, the first interactive item comprises start button **36**B, and target point **120** is positioned on the start button **36**. Returning to step **106**, if the target point is not in proximity to any of the interactive items, then the method continues with step **102**.

In a response step **110**, computer **26** opens (i.e., presents) one or more second interactive items **36** in proximity to the first interactive item, and the method ends. Computer **26** can the select one of the second interactive items responsively to the user adjusting the gaze direction so that the target point is in proximity to the one of the second interactive items.

While the configuration of the interactive user interface in FIGS. **6**A-**6**B **36** show computer **26** presenting the second interactive items **36** radially outward from the first interactive item in a concentric pattern, other configurations are considered to be within the spirit and scope of the present invention. For example, computer **26** may present the first and second interactive items as nested rectangular icons.

In some embodiments, computer **26** can open the one or more second interactive items responsively to a gesture performed by a limb (e.g., finger **30** or hand **31**). For example, after associating the target point with the first interactive item, if the computer receives a sequence of 3D maps indicating that the user is moving hand **30** in a swiping motion (i.e., along a plane comprising X-axis **40** and Y-axis **42**), computer **26** can responsively open the one or more second interactive items.

In the configuration shown in FIG. **6**A, computer **26** presents a menu area **122** of second interactive items comprising tabs **36**C (arranged in a concentric pattern) for various applications that are available on the computer. In some embodiments, computer **26** opens menu area **122** in response to the user gazing at target point **120**, without necessarily requiring the user to make any sort of physical gesture or voice command.

In FIG. **6**B, user **22** has directed his line of sight **92** toward a MAIL tab **36**C, thus causing computer **26** to automatically open the user's electronic mail (e-mail) inbox, which contains listings **36**D of incoming mail items. As the user scans his gaze up and down over listings **36**D and brings line of sight **92** to rest on a particular listing **36**D, the computer automatically presents a preview **124** of the content of the mail item represented by the listing. The user may preview other listings **36**D by moving his line of sight **92** up or down, or may return to the main menu by directing his line of sight **92** to a BACK button **36**E. All of these actions may be performed by eye movement alone.

To open a selected mail item, the user may input a command to computer **26** by another modality. For example, the user may say "OPEN," or may make an opening hand gesture while gazing at the listing **36**D corresponding to the mail item. Sensing device **24** detects the audio input or the 3D motion made by the user and inputs the appropriate command to computer **26**. As a result, the screen shown in FIG. **6**C opens, presenting full content **126** of the selected mail item. The user may return to the inbox from this point by moving his gaze to a "BACK TO INBOX" area **36**F, or may return to the main menu using BACK button **36**E, as mentioned above.

When selecting a given interactive item **36**, computer **26** can convey visual feedback to the user indicating the selection (i.e., before performing an action such as presenting full content **126**). Examples of visual feedback include changing the size and/or appearance of the selected item, or highlighting the selected item by surrounding the selected item with a border. Conveying visual feedback enables user **22** to focus his gaze in the vicinity of the target point, thereby enhancing the user experience. For example, when the user selects start button **36**B, computer **26** can convey visual feedback via the start button, thereby directing the user to maintain his gaze in proximity to where the computer presents tabs **36**C in menu area **122**.

Gaze Related Pointing Gestures

As described in detail hereinbelow, user **22** points to a given interactive item **36** while performing the Point-Select, Point-Touch and Point-Context gestures. In embodiments of the present invention, computer **26** identifies where the user is pointing by defining a line between a first point and a second point of the user, and identifying where (i.e., target point **120**) the line intersects display **28**.

FIG. **7** is a flow diagram that schematically illustrates a method of detecting a gaze related pointing gesture, in accordance with an embodiment of the present invention, FIG. **8** is a schematic pictorial illustration of user **22** performing a Point-Select gesture to select a first given interactive item **36** presented on display **28**, in accordance with an embodiment of the present invention, and FIG. **9** is a schematic pictorial illustration of user **22** performing a Point-Touch gesture to manipulate a second given interactive item **36** presented on display **28**, in accordance with an embodiment of the present invention. Examples of the first interactive items that user **22** can select using the Point-Select gesture include icons **36**G, and examples of the second interactive items that user **22** can manipulate using the Point-Touch gesture include icons **36**G, a vertical scroll box **36**H and a horizontal scroll box **36**I.

The Point-Select gesture comprises user **22** pointing a given finger **30** (typically an index finger) toward the first given interactive item **36**, moving finger **30** (typically along Z-axis **44**) toward the first given interactive item, stopping or slowing down the finger, and then pulling the finger back to the user. The Point-Touch gesture comprises user **22** pointing finger toward a second given interactive item **36**, moving finger **30** (typically along Z-axis **44**) toward the second given interactive item, pausing the finger, and then moving the hand along a plane comprising X-axis **40** and Y-axis **42**.

In a first receive step **130**, a computerized system such as computer **26** receives a first sequence of 3D maps that include user **22**, and in an extraction step **132**, the computer segments the received 3D maps to extract 3D coordinates of the first point and the second point of the user. While FIGS. **8** and **9** show the first point comprising eye **34** and the second point comprising a tip of (index) finger **30**, different first points and second points are considered to be within the spirit and scope of the present invention. For example, the first point may comprise a point between eyes **34**, or some other point on the user's face, and the second point may comprise any point on

any of the user's limbs. For different postures of user **22**, computer **26** may use different calculations to identify the first and the second points.

In a first comparison step **134**, if the first sequence of 3D maps indicates that the user is moving the second point with respect to (typically toward) display **28**, then in a second receive step **136**, computer **26** receives and segments a second sequence of 3D maps indicating a deceleration of the second point. In a calculation step **138**, computer **26** calculates a line segment **160** that intersects the first point and the second point, and in an identification step **140**, the computer extends line segment **160** to display **28**, and identifies target point **120** where line segment **160** intersects the display.

In some embodiments, user **22** may not be looking at the target point. For example, computer **26** may determine (i.e., using the gaze detection embodiments described supra) that the user is directing his gaze at a first given interactive item **36** that is presented on the left side of display **28**, but step **140** identifies that the user is pointing to a second given interactive **36** that is positioned on the right side of display **28**. In this instance, the computer can be configured to select the second interactive item, even though the user is directing his gaze toward the first interactive item.

In a second comparison step **142**, if computer **26** is presenting a given interactive item **36** in proximity to target point **120**, then the computer engages the given interactive item **36** in proximity to the target point, and receives and segments a third sequence of 3D maps in a third receive step **144**. In a third comparison step **146**, if the third sequence of 3D maps indicates that the user is moving the second point away from the display (i.e., toward user **22**), then in a selection step **148**, computer **26** engages (i.e., selects) the given interactive item presented in proximity to target point **120**, and the method ends. In the example shown in FIG. **8**, user **22** can select the icon "A" using the Point-Select gesture comprising pointing finger **30** at the icon "A", and pulling the finger back.

In some embodiments engaging the given interactive item may comprise performing an action associated with the given interactive item. For example, the given interactive item may comprise a given icon **36**G for a movie, and engaging the given icon **36**G comprises executing an application to view the movie.

Alternatively, while pointing at the icon "A", user **22** can issue a vocal command such as "Select". Upon receiving the vocal command from microphone **54**, computer **26** can perform an operation associated with the given interactive item presented in proximity to target point **120** (e.g., the icon "A").

In some embodiments, computer **26** may select, using the Point-Select gesture, a given interactive item **36**, even if the gaze detection embodiments described supra indicate that user is not looking at the given interactive item. In a first example, if computer **26** detects that the user has not changed his gaze direction for a specified time period, and computer **26** identifies a Point-Select gesture while the received 2D image and 3D map do not indicate a gaze direction (e.g., hand **31** is blocking eye **34**), then computer **26** may "override" gaze detection and respond to the Point-Select gesture.

In a second example, computer **26** may respond to a Point-Select gesture, if the finger motion gesture is "significant" (i.e., having at least a first specified difference) and if target point **120** is at least a second specified distance from the interactive item presented in the direction of the gaze. For example, if the user is looking at a first interactive item **36** presented on a left side of display **28**, and performs a Point-Select gesture to a second interactive item **36** presented on an right side of the display, computer **26** may be configured to engage the second interactive item.

If the third sequence of 3D maps do not indicate that the user is moving the second point away from display **28**, then in a fourth comparison step **150**, computer **26** analyzes the third sequence of 3D maps to determine if user **22** is moving the second point along an X-Y plane comprising X-axis **40** and Y-axis **42**. If the third sequence of 3D maps indicates that the user is moving the second point along the X-Y plane, then in a repositioning step **152**, computer **26** repositions the interactive item positioned in proximity to target point **120** responsively to the motion of the second point, and the method ends. In the example shown in FIG. **9**, user **22** can move vertical scroll box **36**H using the Point-Touch gesture comprising pointing finger **30** at the vertical scroll box, and moving the finger up or down.

If the third sequence of 3D maps do not indicate that the user is moving the second point along the X-Y plane, then in a fifth comparison step **154**, computer **26** analyzes the third sequence of 3D maps to determine if user **22** is holding the second point relatively steady for at least a specific time period (e.g., two second). If the third sequence of 3D maps indicates that the user is holding the second point relatively steady for at least the specific time period, then in a presentation step **156**, computer **26** performs a hold operation on the for the interactive item positioned in proximity to target point **120**. For example, the hold operation may comprise presenting context information for the interactive item positioned in proximity to target point **120**. Examples of context information include properties of the interactive item, or options available to the user. For example, computer **26** can present a message "Pull finger back to select, or move finger horizontally or vertically to reposition the item." Another example of the hold operation may comprise computer **26** deleting the interactive item positioned in proximity to target point **120**.

Returning to step **154**, if the third sequence of 3D maps does not indicate that the user held the second point relatively steady for the specific time period, then the method continues with step **130**. Likewise, if computer **26** is not presenting a given interactive item **36** in proximity to target point **120** in step **142**, or if the first sequence of 3D maps does not indicate that user **22** is moving the second point toward display **28** in step **134**, then the method also continues with step **130**.

In some embodiments, computer **26** can control a function of the device coupled to the computer in response to the gestures described herein. For example, if the user performs a Point-Touch gesture in the up direction while pointing toward a speaker coupled to the computer, the computer can raise the volume level of the speaker. Similarly, if the user performs a Point-Touch gesture in the down direction while pointing toward the speaker, the computer can lower the volume level of the speaker.

While the gestures described in the flow diagram of FIG. **7** include the Point-Select and the Point-Touch gestures, other pointing gestures that identify target point **120** via line segment **160** (i.e. intersecting the first and second points) are considered to be within the spirit and scope of the present invention. Additionally or alternatively, the Point-Select and the Point-Touch gestures may be used in conjunction with the gaze detection embodiments described supra.

For example, computer **26** can be configured to select (i.e., Point-Select) a given interactive item **36** upon identifying target position **120** based on the user's gaze, the first sequence of 3D maps indicating user **28** is moving finger **30** toward display **28**, the second sequence of 3D maps indicating that the user is decelerating the finger, and the third sequence of 3D maps indicating that the user is moving finger **30** away from the display. Similarly, computer **26** can be configured to responsively reposition (i.e., Point-Touch) a given interactive

item **36** upon identifying target position **120** based on the user's gaze, the first sequence of 3D maps indicating user **28** is moving finger **30** toward display **28**, the second sequence of 3D maps indicating that the user is decelerating the finger, and the third sequence of 3D maps indicating that the user is moving finger **30** along the plane comprising X-axis **40** and Y-axis **42**.

FIG. **10** is a schematic pictorial illustration showing an alternative Point-Select gesture, also referred to herein as a Trigger gesture in accordance with an embodiment of the present invention. After pointing index finger **30** at a given interactive item **36**, user **22** can select the given interactive item by raising or folding a thumb **170** (as indicated by an arrow **172**), where folding the thumb completes the gestures (i.e., instead of moving the finger toward the given interactive item). Alternatively, computer **26** can select the given icon upon user **22** pointing finger **30** at the given interactive item, and issuing an audio command to select the given interactive item (e.g., the user utters the word "open").

Additionally or alternatively, while the example shown in FIGS. **8-9** show the first point positioned on head **32**, other first points are considered to be within the spirit and scope of the present invention. In the example shown in FIG. **10**, the first point comprises a knuckle of index finger **30**, and line segment **160** intersects the fingertip and the knuckle.

Non-Tactile User Interface Calibration

In operation, non-tactile user interface **20** is typically used by more than one user, and each user may point differently at the same given interactive item **36** presented on display **28**. In some embodiments, using elements of the non-tactile user interface (e.g., icons **30**), computer **26** can calculate and store a calibration coefficient for each user of the non-tactile user interface.

FIG. **11** is a schematic pictorial illustration of user **22** pointing finger **30** at a given icon **36**G, in accordance with an embodiment of the present invention. As shown in the Figure, based the user's positioning of finger **30**, line segment **160** points just below the icon "A". Therefore, upon identifying user **22**, computer **26** can apply an appropriate calibration coefficient when detecting gestures performed by the user.

For example, if non-tactile user interface **20** is inactive for a specified period of time, computer **26** may "lock" the non-tactile user interface, and present a single unlock icon **36**G that user **22** can point to in order to unlock the user interface. As the user points at the unlock icon, computer **26** can identify target point **120** and calculate a calibration coefficient for the user (i.e., based on a proximity of the target point to the unlock icon).

Additionally or alternatively, there may be instances when user **22** performs a gesture, target point **120** is located between two or more icons **36**G, and computer cannot identify which of the icons the user is pointing to. In instances when computer **26** cannot identify which icon **36**G is being pointed to, the computer can identify a subset of the icons that are in proximity to target point **120**, present the identified icons in a larger size, and prompt the user to point again.

FIG. **12**A is a schematic pictorial illustration of user **22** positioning finger **30** to point at one of icons **36**G, in accordance with an embodiment of the present invention. As shown in the figure, target point **120** is located between the icons "A", "B", "E", and "F".

FIG. **12**B is a schematic pictorial illustration of user **22** positioning finger **30** to point at a given icon **36**G, where computer **26** presents the icons in a larger size, in accordance with an embodiment of the present invention. Continuing the example shown in FIG. **12**A, computer **26** presents the icons "A", "B", "E", and "F" in a larger size, and user **22** positions finger **30** to point at the icon "A".

In a similar manner, user interface **20** can calibrate user **22**, when the user selects a given icon **36**G via gaze detection derived from a 2D image and/or a 3D map, as described supra. In a first, computer **26** may "lock" the non-tactile user interface, and present a single unlock icon **36**G that user **22** can look at to in order to unlock the user interface. As the user gazes at the unlock icon, computer **26** can identify target point **120** and calculate a calibration coefficient for the user (i.e., based on a proximity of the target point to the unlock icon).

In a second example, if the user is gazing at a given interactive item **36**G for a specific period of time, computer **26** may present context information for the given interactive item in an interactive item **36** (e.g., a pop-up dialog box). If the user gazes in proximity to the interactive item presenting the context information, computer **26** can calculate a calibration coefficient based on the proximity of the target point to the context information.

Middleware

As described supra, middleware for extracting higher-level information from the 3D maps and gaze information may run on processor **50** and/or CPU **68**, and CPU **68** may execute application programs which drive user interface **72** based on information provided by the middleware, typically via an API.

The following are examples of middleware primitives that computer **26** can be use to extract information from 3D maps received from device **24**:

- InteractStark): Identifies a beginning of an interaction with user interface **20**.
- InteractHover(Pos2D, Radius): Identifies current target point **120** on display **28** (i.e., coordinates on the display where user **22** is pointing finger **30**). The Pos2D parameter references a (two dimensional) location on display **28**.
- InteractPointNew(Pos2D, Radius): Identifies target point **120** when user **22** performs a Point-Select gesture.
- InteractPointUpdate(Pos2D, Radius): Updates target point **120** as user **22** moves finger **30** along the X-Y plane while performing the Point-Touch gesture.
- InteractEnd(Pos2D, Radius): Identifies when user **22** moves finger **30** outside field of view **94**.

Using the middleware functions described supra, computer **26** can identify the following "stages" as user **22** performs the Point-Select and the Point-Touch gestures:

1. Finger Identification. During the Finger Identification stage, the middleware identifies and starts tracking finger **30**. The middleware tracks the finger, and identifies any intent in the finger's motion toward the display. This enables a user interface **20** to differentiate a pointing gesture from other casual hand movements.
2. Gesture Identification. To reach the Gesture Identification stage, two conditions have typically been met: (a) User **22** moved finger **30** toward display **28**, and (b) Target point **120** is within the boundaries of the display. In some configurations, target point **120** can be "located" slightly outside display **28**. For example, since a valid Point-Touch gesture may include dragging a given interactive item **36** from "outside" display **28** (e.g., backspace, alt-tab etc.).

Upon entering the second stage, the middleware conveys an InteractStart( ) event. Upon conveying the InteractStart( ) event, the middleware is tracking finger **30** con-

stantly and looking for a change in direction which defines the target point (see stage 3 below). While the finger is moving, the middleware is conveying InteractHover(Pos2D, Radius) events, which enables user interface **20** to detect where user **22** is currently pointing finger **30**.

3. Interaction Point. When user **20** stops finger **30**, or moves finger **30** away from display **28**, the Interaction Point stage is reached. Based on a location of the second point, the middleware calculates target point **120** by connecting (for example) eyes **30** and the finger point to create line segment **160**, and extends the line segment to reach the display. The middleware then conveys the event InteractPointNew(Pos2D, Radius) which identifies target point **120**, thereby enabling user interface **20** to select the intended interactive item **36**.
4. Interaction Ends. User **22** moving finger **30** away from display **28** and outside field of view **94** indicates completion of a gesture (e.g., the Point-Select and the Point-Touch gestures described supra), thereby disengaging the user from user interface **20**. To re-engage user interface **20**, user **22** can reposition finger **30** within field of view **94**, thereby entering the finger identified stage described hereinabove.

At a higher level of abstraction, the middleware primitives described supra can be combined to define the following primitives that are similar to existing touch screen primitives that are known in the art:

Activate: The Point-Select gesture that engages (i.e., "clicking") a given interactive item **36**. The Activate primitive can be used to activate an application, pressing button, or follow a hyperlink.

Pan: The Point-Touch gesture that moves a given interactive item **36** in any direction on the X-Y plane. In operation, some applications may be configured to react to movements on either the X-axis or the Y-axis.

Context: As described supra, the Point-Hold gesture comprising user **22** moving finger **30** toward a given interactive item **36**, and the holding the finger relatively steady for a specific timer period (i.e., similar to positioning a mouse over an Item and pressing the right mouse button in a Microsoft Windows™ environment). In response to the Point-Hold gesture, computer **26** can convey feedback to the user indicating what to do next (e.g., dragging and/or drop a given icon **36**G), or information on the given interactive item (e.g., movie plot summary as described supra).

In operation, the following events can be conveyed by the middleware in order to allow applications executing on computer **26** to be developed at a higher abstraction level which reflects user experience (UX) language primitives:

Start(Pos2D, Radius): Initiate interaction with user interface **20** in a manner similar to the InteractStart( ) event but at a higher level of abstraction.

Activate(Pos2D, Radius): Activate one of the gesture primitives described hereinbelow.

Pan(Pos2D, radius): User **22** initiates a Point-Touch gesture.

PanEnd(pos2D, Radius): User **22** completes (i.e., disengages from) a Point-Touch pan gesture (e.g., by moving finger **30** back away from the display).

Context(Point2D, Radius): User **22** initiates a Point-Hold gesture. In operation, user **22** may transition directly from the Point-Hold gesture to the Point-Touch gesture.

ContextEnd( ) User **22** disengages from user interface **20** upon completing the Point-Hold gesture (i.e., without transitioning to the Point-Touch gesture).

The combined 3D mapping and gaze direction information provided by sensing device **24** may be used in various ways to enhance the quality and user experience of 3D graphical rendering and 3D image presentation. For example, in an interactive game in which the user moves through a scene, scenery on display **28** may be brought into focus in the direction of the line of sight **92**. Scenery in other areas (such as distant scenery when the user's point of view is on a nearby item, or vice versa) may be intentionally blurred, in order to simulate actual depth accommodation and/or to save bandwidth.

As another example, a game application running on computer **26** may be programmed to change a "story line" of the game depending on the user's gaze direction. For instance, the computer may surprise the user by presenting items (such as "enemy" characters) suddenly in areas of the display where the user is not looking at a given moment. The methods of gaze-directed pointing and selection that were described above in reference to FIGS. **6**A-**6**C may likewise be applied in selecting interactive items **36** and "aiming" them (such as pointing a weapon at a target) in games and other "virtual worlds."

Depth and gaze information collected by device **24** can be used in enhancing the capabilities and user experience of 3D displays, particularly autostereoscopic displays. Such displays operate by presenting different images to the user's right and left eyes, but generally can be viewed only from a limited range of positions. By tracking the user's head location and gaze direction, device **24** may be able to direct the display to modify any images it presents so that they can be viewed over a larger range of positions and show different angular views of the items presented on the display. The parallax applied to near-field items **36** that are presented on the autostereoscopic display (or other 3D display) can be modified depending on the distance of the user's head from the display, in order to enhance realism and reduce visual discomfort that some users may feel in this environment.

This sort of 3D display can also be driven to interact with the user's 3D gestures. For example, based on a known location of the user, as well as the user's gaze direction, computer **26** can drive the 3D display to display virtual items in space at locations where the user can "touch" them. The user can then manipulate and interact with the virtual items by moving his hands (or other body parts) in the locations of the items in 3D space. Device **24** senses the user's gestures and provides appropriate input to computer **26**, so that the computer can move or otherwise modify the items in response to the user interactions. This sort of interaction model also enables the user to reach for and interact with a given on-screen object that is located "behind" another object in the virtual space created by the display.

In an additional example, computer **26** may present content comprising multiple interactive items **36** (e.g., characters in a game) on display **28**. In some embodiments, computer **26** can identify a region (i.e., an area around target point **12**) on the display in the direction of the user's gaze, and present the content within the identified region as "in focus" (i.e., present clearly), and the content outside the identified region as "out of focus" (i.e. present blurred). In an alternative embodiment, each of the multiple interactive items may have an associated depth value, and as the user gazes toward a given interactive item, the computer can simulate a 3D environment by presenting the interactive items whose associated depth values are in accordance with the depth value if the given interactive item as "in focus" (i.e., present clearly), and presenting the

interactive items whose associated depth values are not in accordance with the depth value if the given interactive item as "out of focus" (i.e. present blurred).

In a further example, while playing a game, user **22** can select a weapon by pointing at a weapon presented as an interactive icon on the display. If the selected weapon is a gun, the user can "aim" the gun using the Point-Touch gesture, and "shoot" the gun using the Point-Select gesture or the Trigger gestures described supra. Alternatively, if the selected weapon is a sword, then the user can manipulate the sword in three dimensions (i.e., along the X-Y plane and the Z-axis) by using a combination of the Point-Select and the Point-Touch gestures.

The above applications are just a few examples of how mixed modality user interfaces can be used to enhance system capabilities and user experience, and other, similar applications are considered to be within the scope of the present invention. As another example, the capabilities of sensing device **24** may be used in gauging user interest in content such as a Web site or video program, depending on the on-screen object on which the user fixed his gaze, as well as whether the user was looking at the display at all while a certain program (such as a commercial) was presented. These capabilities may similarly be used in extracting user interest profiles, as described, for example, in U.S. patent application Ser. No. 13/295,106, filed Nov. 14, 2011, which is incorporated herein by reference.

It will be appreciated that the embodiments described above are cited by way of example, and that the present invention is not limited to what has been particularly shown and described hereinabove. Rather, the scope of the present invention includes both combinations and subcombinations of the various features described hereinabove, as well as variations and modifications thereof which would occur to persons skilled in the art upon reading the foregoing description and which are not disclosed in the prior art.

The invention claimed is:

1. A method, comprising:

receiving a first sequence of three-dimensional (3D) maps over time of at least a part of a hand of a user of a computerized system;

segmenting the 3D maps in order to extract 3D coordinates of a first point located on a knuckle of the finger on the hand and a second point at a tip of the finger;

calculating a line segment that extends through the first point and the second point to a display that is coupled to the computerized system;

identifying a target point where the line segment intersects the display; and

in response to a motion of the hand indicated by the 3D maps, engaging an interactive item presented on the display in proximity to the target point.

2. The method according to claim **1**, wherein engaging the interactive item comprises performing an operation associated with the interactive item upon receiving a vocal command from the user.

3. The method according to claim **1**, wherein the motion of the hand comprises a motion toward the display.

4. The method according to claim **3**, wherein engaging the interactive item comprises performing an action associated with the interactive item upon receiving a second sequence of 3D maps indicating a deceleration of the motion of the second point toward the display, and receiving a third sequence of 3D maps indicating a motion of the second point away from the display.

5. The method according to claim **3**, wherein engaging the interactive item comprises repositioning the interactive item on the display responsively upon receiving a second sequence of 3D maps indicating a deceleration of the motion of the second point toward the display, and receiving a third sequence of 3D maps indicating a motion of the second point along a plane comprising vertical and horizontal axes.

6. The method according to claim **3**, wherein engaging the interactive item comprises performing a hold operation on the interactive item responsively upon receiving a second sequence of 3D maps indicating a deceleration of the motion of the second point toward the display, and receiving a third sequence of 3D maps indicating that the user is holding the second point relatively steady for a specific time period.

7. The method according to claim **6**, wherein the hold operation comprises presenting context information for the interactive item.

8. The method according to claim **6**, wherein the hold operation comprises deleting the interactive item from the display.

9. The method according to claim **3**, wherein the finger comprises an index finger, and wherein engaging the interactive item comprises selecting the interactive item upon receiving a second sequence of 3D maps indicating a deceleration of the motion of the second point toward the display, and receiving a third sequence of 3D maps indicating the user folding a thumb toward the index finger while pointing the index finger at the interactive item.

10. The method according to claim **1**, and comprising calculating a calibration coefficient based on the proximity of the interactive item to the target point.

11. The method according to claim **1**, and comprising identifying the line segment intersecting a device coupled to the computer, and controlling a function of the device responsively to a motion of the second point.

12. An apparatus, comprising:

a display; and

a computer coupled to the display and configured to receive and segment a first sequence of three-dimensional (3D) maps over time of at least a part of a hand of a user of a computerized system in order to extract 3D coordinates of a first point located on a knuckle of the finger on the hand and a second point at a tip of the finger, to calculate a line segment that extends through the first point and the second point to the display, to identify a target point where the line segment intersects the display, and in response to a motion of the hand indicated by the 3D maps, to engage an interactive item presented on the display in proximity to the target point.

13. The apparatus according to claim **12**, wherein the computer is configured to engage the interactive item by performing an operation associated with the interactive item upon receiving a vocal command from the user.

14. The apparatus according to claim **12**, wherein the motion of the hand comprises a motion toward the display.

15. The apparatus according to claim **14**, wherein the computer is configured to engage the interactive item by performing an action associated with the interactive item upon receiving a second sequence of 3D maps indicating a deceleration of the motion of the second point toward the display, and receiving a third sequence of 3D maps indicating a motion of the second point away from the display.

16. The apparatus according to claim **14**, wherein the computer is configured to engage the interactive item by repositioning the interactive item on the display responsively upon receiving a second sequence of 3D maps indicating a deceleration of the motion of the second point toward the display,

and receiving a third sequence of 3D maps indicating a motion of the second point along a plane comprising vertical and horizontal axes.

17. The apparatus according to claim **14**, wherein the computer is configured to engage the interactive item by performing a hold operation on the interactive item responsively upon receiving a second sequence of 3D maps indicating a deceleration of the motion of the second point toward the display, and receiving a third sequence of 3D maps indicating that the user is holding the second point relatively steady for a specific time period.

18. The apparatus according to claim **17**, wherein the computer is configured to perform the hold operation by presenting context information for the interactive item.

19. The apparatus according to claim **17**, wherein the computer is configured to perform the hold operation by deleting the interactive item from the display.

20. The apparatus according to claim **14**, wherein the finger comprises an index finger, and wherein the computer is configured to engage the interactive item by selecting the interactive item upon receiving a second sequence of 3D maps indicating a deceleration of the motion of the second point toward the display, and receiving a third sequence of 3D maps indicating the user folding a thumb toward the index finger while pointing the index finger at the interactive item.

21. The apparatus according to claim **12**, wherein the computer is configured to calculate a calibration coefficient based on the proximity of the interactive item to the target point.

22. The apparatus according to claim **12**, wherein the computer is configured to identify the line segment intersecting a device coupled to the computer, and to control a function of the device responsively to a motion of the second point.

23. A computer software product comprising a non-transitory computer-readable medium, in which program instructions are stored, which instructions, when read by a computer, which is coupled to a display, cause the computer to receive and segment a first sequence of three-dimensional (3D) maps over time of at least a part of a hand of a user of a computerized system in order to extract 3D coordinates of a first point located on a knuckle of the finger on the hand and a second point at a tip of the finger, to calculate a line segment that extends through the first point and the second point to the display, to identify a target point where the line segment intersects the display, and in response to a motion of the hand indicated by the 3D maps, to engage an interactive item presented on the display in proximity to the target point.

* * * * *